Journal of Physical Chemistry & Biophysics

Nikolopoulos et al., J Phys Chem Biophys 2014, 4:2
DOI: 10.4172/2161-0398.1000145

Research Article Open Access

Factors Affecting Indoor Radon Concentrations of Greek Dwellings through Multivariate Statistics

Dimitrios Nikolopoulos[1*], Sofia Kottou[2], Anna Louizi[2], Ermioni Petraki[1,3], Efstratios Vogiannis[4] and Panayiotis H Yannakopoulos[1]

[1]*Department of Computer Electronic Engineering, TEI of Piraeus, Greece, Petrou Ralli & Thivon 250, 122 44,Aigaleo,Athens, Greece*
[2]*Medical Physics Department, Medical School, University of Athes, Mikras Asias 75, 11527, Goudi, Athens, Greece*
[3]*Department of Engineering and Design, Brunel University, Kingston Lane, Uxbridge, Middlesex UB8 3PH, London, UK*
[4]*Evangeliki Model School of Smyrna,Lesvou 4, 17123, N. Smirni-Greece*

Abstract

A large scale nationwide radon survey was conducted in Greek dwellings between 1994 and 2000. Twelve hundred passive CR-39 detectors were distributed and collected along with 963 filled in questionnaires. These were rechecked during 2012-13 to evaluate factors that affect indoor radon concentrations, such as i) area, ii) building level-floor, iii) ground type, iv) basement, v) building type, vi) construction year, vii) building walls contact, viii) wall materials, ix) floor materials. The questionnaires were prepared by the research team according to international standards. One-way and multivariate statistical methods were applied for the analysis: i) Linear Regression Analysis, ii) One way or multiway ANOVA, iii) General MANOVA, iv) Stepwise Regression Analysis, v) Principal Components Analysis. Results revealed that approximately 0.1% of the dwellings exhibited outlier radon concentrations. Noteworthy statistical correlations were detected between indoor radon concentration and the factors: "building level-floor" and "wall materials". Results of current work strengthened these considerations and provided weak evidence for the correlation of factors like "building type", "construction year" and "floor materials" with radon concentrations. Minor association was detected with the factor "building walls contact". Significant differences were detected in results produced by some of the applied statistical methods.

Introduction

Natural environmental radiation depends on local geology and hence, variations are addressed in human radiation exposure due to cosmic and terrestrial radiation [1]. ^{238}U and ^{232}Th are two natural parent isotopes which are present in soil and contribute significantly to natural terrestrial radioactivity. Radon ^{222}Rn is a radioactive noble gas and originates from ^{238}U. ^{220}Rn originates from ^{232}Th and ^{219}Rn from ^{235}U. ^{222}Rn, ^{220}Rn, ^{219}Rn are the primary sources of radon in soil, with ^{222}Rn being dominant in rocks, soil, building materials, underground and surface waters [2] and set to be the most hazardous radionuclide. Radon (^{222}Rn) and its short-lived progeny (^{218}Po, ^{214}Po, ^{214}Bi, ^{214}Pb) are attached in dust and in water droplets creating radioactive aerosols, that inhaled via breathing and enter human lungs. Radon enters buildings through gaps around pipes or cables and through cracks in floors [3]. Primary studies have shown that radon is the second most dangerous cause of lung cancer after smoking. This happens when alpha particles emitted from radon progeny damage pulmonary epithelium [4-7]. Many studies have been made for the measurement of indoor radon concentrations in several countries [8-19]. Over the years, in Greece, indoor radon concentrations measurements, to our knowledge, are as follows: several small–scale [20-24], two middle–scale [3,25] and one large-scale [26].

Under the National Strategic Reference Framework (NSRF) "Thales" project of the Technological Education Institute of Piraeus and extending the aforementioned large scale radon survey; this paper addresses the grade of severity with witch factors influence indoor radon concentration levels. Similar studies in other countries have shown that indoor radon concentrations are higher at lower floor levels [27-30]. Moreover, recent studies indicate that radon emanation from building materials contributes significantly in indoor radon concentration in dwellings [31-33]. Results of current work may strengthen these considerations and additionally provide evidence for lack of correlation of other factors such as the "building walls contact" and "construction year" with radon concentrations.

Materials and Methods

A thorough investigation was performed on whether nine factors: i) area, ii) building level-floor, iii) ground type, iv) basement, v) building type, vi) construction year, vii) building walls contact, viii) wall materials, ix) floor materials, may affect indoor radon concentration independently or jointly. These factors have been recorded on 963 filled questionnaires of the greater large-scale survey in Greece [26]. A multivariate statistical analysis, based on i) Linear Regression Analysis, ii) One way or multiway ANOVA, iii) General MANOVA, iv) Stepwise Regression Analysis and v) Principal Components Analysis methods, was implemented on the questionnaire data. It is noted that these data were dispersed across Greece.

Linear regression analysis

In linear regression analysis, a straight line is fitted through a set of points-observations in such a way that the sum of squared residuals is minimal [34]. In multiple linear regression the dependent variable can be written in terms of a linear combination of the independent variables. The regression equation describes the correlation of the mean value of a variable-y with specific values of x-variables used to predict y.

***Corresponding author:** Dimitrios Nikolopoulos, Department of Computer Electronic Engineering, TEI of Piraeus, Greece, Petrou Ralli & Thivon 250, 122 44, Aigaleo, Athens, Greece, Tel: +0030-210-5381110, Mobile +0030-6977-208318; Fax: +0030-210-5381436; E-mail: dniko@teipir.gr, NikolopoulosDimitrios@gmail.com

Received April 2, 2014; **Accepted** May 22, 2014; **Published** May 24, 2014

Citation: Nikolopoulos D, Kottou S, Louizi A, Petraki E, Vogiannis E, et al. (2014) Factors Affecting Indoor Radon Concentrations of Greek Dwellings through Multivariate Statistics. J Phys Chem Biophys 4: 145. doi:10.4172/2161-0398.1000145

Suppose that $(x_1,y_1),(x_2,y_2),...,(x_n,y_n)$ are the realisations of random variable pairs $(X_1,Y_1),(X_2,Y_2),...,(X_n,Y_n)$, then the linear regression equation expresses the mean of Y as a straight-line function of X and could be represented as

$$E(Y_i)=\beta_0+\beta_1 \cdot X_i \quad (2.1.1)$$

or $E(Y_i)=\beta_0+\beta_{i1}X_1+\beta_{i2}X_2+\ldots+\beta_{ip}X_p$ for p independent-predictor variables.

$E(Y_i)$ states the mean expected value and i points the population. The estimated/fitted model is then:

$$\hat{Y}=\beta_0+\beta_1 \cdot X \quad (2.1.2)$$

From (2.1.2), the estimated/fitted values for each of the n observations are

$$\hat{Y}_i=\beta_0+\beta_1 \cdot X_i \quad (2.1.3)$$

where $i=1,\ldots,n$ is the consecutive number of the population. From (2.1.2) and (2.1.3) the, so called, observed error or fitted residual is calculated as:

$$e_i=Y_i-\hat{Y}_i \quad (2.1.4)$$

Equation (2.1.4) calculates the estimated error of the i-th observation in the sample. From (2.1.4) the sum of squared observed errors (SSE) equals

$$SSE=\sum(Y_i-\hat{Y}_i)^2=\sum e_i^2 \quad (2.1.5)$$

for all observations in a sample of size n. The mean square error (MSE) equals then

$$MSE=\frac{SSE}{n-2}=\frac{\sum e_i^2}{n-2} \quad (2.1.6)$$

("n-2" should be substituted by "n-p-1" when there are p predictor-independent variables) and this is the sample variance of error. The residual standard error is then calculated as

$$\hat{\sigma}=\sqrt{MSE} \quad (2.1.7)$$

and σ^2 should be the constant error variance, otherwise the confidence intervals will be misleading.

$$\text{As } \underset{\text{Total deviation}}{Y_i-\bar{Y}} = \underset{\text{Deviation due to the regression}}{\hat{Y}_i-\bar{Y}} + \underset{\text{Deviation due to the error}}{e_i} \quad (2.1.8)$$

then, the total sum of squares (SST) equals

$$\underset{SST}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}=\underset{SSR}{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}+\underset{SSE}{\sum_{i=1}^{n}e_i^2} \quad (2.1.9)$$

with $\bar{Y}$ set to be the mean of all observed Y values.

The coefficient of determination

$$R^2=\frac{SSR}{SST}=1-\frac{SSE}{SST} \quad (2.1.10)$$

represents the proportion of variation in Y that is explained by X [35]. Parameters β_0, β_1, ..., β_p (regression coefficients), σ^2 (variance) and R^2 (coefficient of determination) need to be estimated in order to examine if the linear regression model applies to this group of data. However, even if R^2 value is close to zero, this does not mean X and Y have no nonlinear association and polynomial terms should be included to improve the fitting.

One Way or multiway ANOVA: Analysis of variance (ANOVA) approach to regression analysis is considered as the generalization of a t-test to more than two statistical groups. ANOVA is divided in two categories: i) One way, where a single factor exists and ii) two way or multiway, where two or more factors exist. ANOVA is used for distributional assumptions about a set of effects in a model, with ability to extrapolate the inferences to a wider population, improve accounting for system uncertainty and the efficiency of estimation [36]. ANOVA has been implemented as the basic method for the statistical analysis of Radon concentrations in many studies [15,19,37-40]. For the analysis of factors affecting indoor radon concentrations, initially each factor was analysed independently. In such a way a first assumption for the weightiness of the effect of each factor is possible. In the next step, factors affection are no longer estimated independently; instead, factors influence each other and therefore are dependent. The aforementioned method, is called random-effects assumption of the analysis of variance [36]. ANOVA can be implemented only in sampling distributions similar to Gaussian ones, thus it was applied in the log distributions of radon measurements.

In order to construct the ANOVA table the variability in Y variable explained or not with the regression relationship of X variable was measured. This table shows additionally the Mean Square Error (MSE). The overall variation in Y is equal with the sum of regression variation and the error variation (2.1.9).

The ANOVA table (Table 1) involves the following elements:

The sum of squares for total (*SST*) which is the sum of the squared deviations from the overall mean of *Y*.

The sum of squared errors (*SST*) which is the sum of squared

Source of variation	SS	df	MS	F-value
Regression	$SSR=\sum_{i=1}^{n}(Y_i-\bar{Y})^2$	p	$MSR=\frac{SSR}{p}$	$F=\frac{MSR}{MSE}$
Error	$SSE=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$	$n-p-1$	$MSE=\frac{SSE}{n-p-1}$	
Total	$SST=\sum_{i=1}^{n}(Y_i-\bar{Y})^2$	$n-1$	$\frac{SST}{n-1}$	

Table 1: The ANOVA table.

observed errors for the observed data and is a measure of the variation in *Y* which is not explained by the regression (2.1.5).

The sum of squares of the regression (*SST*) defined as the difference between *SST* and *SSE*. This is a measure of the total variation of *Y* that can be explained from the regression with *X* variable.

The mean of square errors (2.1.6) and

The mean square of the regression (*MSR*) which, here, equals with the *SSR* [34].

The F-value, where F(1-α, 1, n-2) is the (1-α) quantile of the F distribution and α is the significant level.

General MANOVA: Multivariate (multiple dependent variables) analysis of variance (MANOVA) is defined as a partition of the sum of squares and the sum of the cross products (SSCP) matrix

$$SSCP=\begin{bmatrix} SS_{11} & SCP_{12} \\ SCP_{21} & SS_{22} \end{bmatrix} \quad (2.1.2.1)$$

into independent Wishart matrices [41]. MANOVA is applied instead of a series of one-at-a-time ANOVAs. In several situations, the power of MANOVA is inferior to ANOVA of one variable at a time, however, MANOVA takes into account the intercorrelations among the dependent variables. Hence, MANOVA is considered more efficient over ANOVA for multivariate data [42].

A MANOVA table (Table 2) includes the following elements:

The sum of squares and cross products for total (*SSCPTO)* which is the sum of squared deviations from the overall mean vector of the Y_i and equals to

$$SSCPTO=\sum_{i=1}^{n}(Y_i-\bar{Y})(Y_i-\bar{Y})^T \quad (2.1.2.2)$$

SSCPTO is considered as a measure of the overall variation in the Y vectors.

The sum of squares and cross-products of the errors (*SSCPE*)

$$SSCPE=\sum_{i=1}^{n}(Y_i-\hat{Y})(Y_i-\hat{Y})^T \quad (2.1.2.3)$$

which is the sum of squared errors (residuals) for the data vectors. *SSCPE* is considered as a measure of the variation in Y that is not explained by the multivariate regression.

The sum of squares and cross-products due to the regression (*SSCPR*) is defined as the difference between *SSCPE* and *SSCPE* :

$$SSCPR=SSCPTO-SSCPE \quad (2.1.2.4)$$

Source of variation	SSCP	df
Regression	$\sum_{i=1}^{n}(Y_i-\bar{Y})(Y_i-\bar{Y})^T$	$p-1$
Error	$\sum_{i=1}^{n}(Y_i-\hat{Y}_i)(Y_i-\hat{Y}_i)^T$	$n-p$
Total	$\sum_{i=1}^{n}(Y_i-\bar{Y})(Y_i-\bar{Y})^T$	$n-1$

Table 2: The MANOVA table.

SSCPR is a measure of the total variation in Y that can be explained by the regression with the predictors [35].

Issues in Multiple Regression: The difficulty with model selection emerges from the fact that for p predictors, there are 2^p different candidate models. With so many possible interactions it can be difficult to find a good model. Model selection methods try to simplify this task. A true model may only depend on a subset of $X_1, ..., X_p$. In other words, in model

$$Y=\beta_0+\beta_1X_1+...+\beta_pX_p+\varepsilon \quad (2.1.3.1)$$

some of the coefficients are zeros. The result will be the disclosure of those predictors with nonzero coefficients, i.e. the "best subset" of all predictors.

R^2 can be used for models with the same number of parameters/ coefficients, otherwise R^2_{adj} should be used. The best model has the biggest R^2_{adj} value.

Selecting p predictors, the Mallows' C_p criterion should be small with a value near to p.

Stepwise Regression Analysis

Stepwise methods are used in several areas of applied statistics. A statistical model can be constructed in two ways, namely (i) forward selection and (ii) backward elimination. Forward selection means that a specific number of variables exist in the beginning and gradually variables are added, one at a time, in optimal way in order to analyse the effect of each variable. Alternatively, with backward elimination, all potential variables exist in the beginning and non-effective variables are subtracted, one at a time, until a desirable stopping point is reached.

Stepwise regression forms a hybrid model between forward selection and backward elimination. More precisely, steps have a forward direction with variable addition, however if a variable is characterized as non-significant, it is removed as in backward elimination [35,43]. In literature stepwise regression has been used for the prediction of mean indoor radon concentrations [44], in the construction of radon maps based in indoor radon measurements and soil geochemical parameters [45] and in risk analysis of factors affecting lung cancer [46].

Principal Components Analysis

Principal components analysis (PCA) is used for the reduction of the number of possible clusters. PCA offers the ability for the identification of patterns within large sets of data [47]. Its significance rely in the occurrence of a relative redundancy in the variables, due to their correlation in the measurement of the same construct. During the analysis of the principal components, eigenvalues represent the relative participation of each factor in presenting the general variability of the sampled data [48]. PCA has several implementations in factors investigation of water quality, in drug development, in cancer detection and in health care [48-53]. PCA has been also used for the investigation of the dependence among variables and for the prediction of relationships among variables [54].

A standardisation of the various data is performed prior to analysis in order to ensure that each variable influences in the same way during the analysis. The standardised variable is the following

$$z_i=\frac{x_i-\mu_i}{\sqrt{\sigma_{ii}}},i=1,\ldots,p \quad (2.3.1)$$

with x_i set to be the original variable, μ_i the mean and Y_i the

variance.

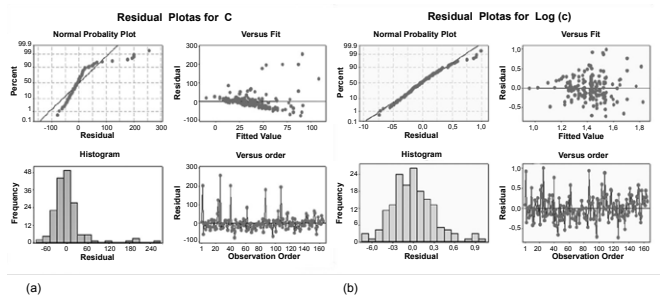
Principal components Y_i are linear combinations of P random variables $X_1, X_2, \ldots, X_p$. If S=S_{ik} is a pxp sample covariance matrix with eigenvalue-eigenvector pairs $(\lambda_1, e_1), (\lambda_2, e_2) \ldots, (\lambda_p, e_p)$, and x set to be the number of principal components, then the i-th principal component sample is given by

$$Y_i = e_i x = e_{il} x_1 + e_{i2} x_2 + \ldots + e_{ip} x_p \quad (2.3.2)$$

where x stands as any observation on the variables $X_1, X_2, \ldots, X_P$ and $i = 1,2,\ldots p$ [55].

Results and Discussion

Figure 1 presents characteristic residual plots calculated from the measured average concentrations of radon (C) and their logarithms (log(C). It is noted that the C values of Figure 1 correspond to time-integration over a year, constitute representative sample for Greece, were derived in accordance to international standards and delineate the radon profile of Greece [26]. In this consensus, Figure 1 is of significance since it may show actual tendencies regarding the randomness or predictability of the employed concentration sample. Indeed, completely randomised responses to normal-distribution either of C or og (C), would exhibit no deterministic normal-distribution's residuals and, hence, be completely described by stochastic processes. The normal probability plots of Figures 1a and 1b indicate, however, that the logarithms of the measured concentration followed normal distribution up to the 95% of og (C) values, namely indicated that$C$ values followed log-normal distribution. This is also evident from the shapes of the frequency distributions of the residuals. The frequency distribution of Fig.1a was clearly log-normal, while simultaneously that of Figure 1b, clearly normal. This is also of significance because all international large-scale radon surveys reported log-normal behaviour of indoor radon concentrations. The reason is rational thus why C values did not follow normal distribution as shown in the corresponding normal probability plot of Figure 1a. Under another view, the residual plots of log (C) versus values fitted to normal-distribution, showed a random pattern for fitted residual values of log (C) above 1.6. It is noted that a residual of 1.6 inlog (C) is consistent with uncertainty $\sigma_C = 39.8\ Bq \cdot m^{-3}$ in predicted C values. This, according to the recording capabilities of the employed dosemeters [56], accompanies high C values, namely C values above the EU action limit of 200 $Bq \cdot m^{-3}$. Moreover, the majority of predicted residuals were below 1.2. This is very significant because this residual range is consistent with concentrations usually addressed, namely between 10-120 $Bq \cdot m^{-3}$. Other factors may affect concentrations in this range and surely the potential factors could not be continuous under the normal distribution. Indeed, the Versus Fits diagram of Figure 1a shows characteristic predictability different from the normal distribution for the residual C range below 75 $Bq \cdot m^{-3}$. On the other hand, the residual versus observation order did not showed tendencies for concentrations up 160 $Bq \cdot m^{-3}$ either in the concentration order (Figure1a) or the order of concentration's logarithm.



(a) (b)

Figure 1: Residual plots of (a) radon concentration and (b) logarithm of radon concentration.

All measurements are shown. Each plot contains (I) percentage of residuals versus residual, (II) frequency distribution of the residuals, (III) residual versus fitted values according to C for (a) and log (C) for (b) and (IV) residual versus observation order.

Table 3 presents the analysed factors, factor levels and level values with their corresponding description. Data of Table 3 were formulated in accordance to the contents of the 963 questionnaires which were filled during the radon survey of Greece. It is noted that these questionnaires were developed in agreement to other national surveys. The majority of factors exhibited 3-4 levels. This is noteworthy in any related analysis, since a multi-level collection of factors can distract results especially for limited number of measurements. Factor (Level-L) was 5-level marking the usual situation of apartment dwellings in big cities of Greece. Nevertheless, this 5-level factor is easily convertible to a lower-level one. Factor (Floor's material- F) was free to fill, so a 6 level collection was finally achieved.

It is well identified that Gauss distribution offers a rigid and justified pathway for statistical analysis. Since concentrations' logarithms followed the distribution of Gauss, log (C) was considered favourable. Hereafter any analysis was conducted only on log (C). Table 4 presents the unusual observations in log (C) values accounting that these followed normal distribution, viz. were treated according to the distribution of Gauss. Leverage points were considered to be those observations corresponding to extreme or outlying values of log (C) in a manner that any lack of neighbouring observations implied that the fitted Gaussian regression model passed close to the particular observation. In specific, leverage points were calculated by moving all points one-by-one up or down and calculating the proportionally constant (leverage) of the change of the corresponding Gaussian fitted value. Outliers were calculated as the observations that presented residuals above 1.5 times the interquartile range. According to Table 4, eight outlier and two leverage residual points were identified. In any case, unusual log (C) values were approximately 0.1% of the total concentration sample size. Therefore, they constituted a negligible part of measurements. Importantly, however, the latter finding indicates that only a small portion (<0.1%) of Greek dwellings presented unusual concentrations. Considering that high unusual concentration extremes may associate with high human radiation burden, this fact implies that indoor radon in Greece may not lie in the international extremes. Emphasis should be stressed also on the fact that outlier data affect any type of fit and should be removed prior to regression analysis, whereas, leverage point may or may not affect. For this reason, all outlier and leverage points were finally removed from the dataset.

Table 5 presents the combinations to define the best subsets from the nine factors of Table 3 for the regression of log (C). As in Table 4, regression was linear to the factors employed in each entry of Table 5. Mallow's C_p (MCP) was calculated for any subset of k, $k \leq p$ of explanatory variables, as $C_p = \frac{SSE_p}{MSE} - (n - 2 \cdot p)$, where SSE_p was the residual sum of squares for the subset model containing p explanatory variables counting the intercept (i.e., the number of parameters in the subset model) and n is the sample size. It should be emphasised that, acceptable models in the sense of minimizing the total bias of predicted values, are those models for which C_p approaches the value p, i.e.,

Factor	Type	Levels	Values	Description
Area-A	fixed	3	1; 2; 3	Agricultural, Suburban, Urban
Level-L	fixed	5	0; 1; 2; 3; 4	Ground Floor,1st,2nd,3rd , >4th floor
Ground Type-G	fixed	3	0; 1; 2	Normal, Sloppy, Over-grounded
Basement-Bas	fixed	3	0; 1; 2	No bas, Full coverage, Semi coverage
Building Type-BTy	fixed	3	0; 1; 2	Detached, Semi-Det, Apartment
Year-Y	fixed	4	0; 1; 2; 3	<1900, 1900-50, 1950-80, >1980
Wall Contact-Con	fixed	4	0; 1; 2; 3	no, 1, 2, 3
Wall's material-W	fixed	4	0; 1; 2; 3	brick/concrete, stone, combin, other
Floor's material-F	fixed	6	0; 1; 2; 3; 4; 5	concr/tile, concr/mos, concr/marble, concr/wood, concr, rock

Table 3: Descriptions of used factors (qualitative)

O	$\log(C)$	Fit	SE-F	Residual	SR
2	2.43459	1.52274	0.13233	0.91185	2.95 R
5	1.15278	1.15278	0.33620	-0.00000	* X
26	2.54306	1.54974	0.15743	0.99332	3.34 R
40	2.46026	1.70575	0.10492	0.75451	2.36 R
45	1.01375	1.68231	0.14050	-0.66856	-2.19 R
56	1.02208	1.02208	0.33620	0.00000	* X
87	2.35829	1.43778	0.08382	0.92051	2.83 R
97	0.72033	1.43709	0.13810	-0.71676	-2.34 R
108	2.39661	1.52606	0.07437	0.87055	2.66 R
126	0.80896	1.54284	0.09904	-0.73388	-2.28 R

Table 4: Unusual Observations for log (C) . O denotes the number of observation, SE-F the standard error of the fit and SR the standardised residual. R denotes observations with large standardized residual. X denotes observations that gave large leverage.

V	R^2	R^2_{adj}	MCP	SE	A	L	G	Bas	B Ty	Y	Con	W	F
1	2.1	1.5	-0.8	0.33511	X								
1	1.1	0.5	0.8	0.33677							X		
2	2.9	1.7	-0.2	0.33465	X						X		
2	2.8	1.6	-0.0	0.33481	X							X	
3	3.7	1.9	0.5	0.33434	X						X	X	
3	3.6	1.8	0.6	0.33446	X	X						X	
4	4.6	2.2	1.1	0.33384	X	X					X	X	
4	4.0	1.6	2.0	0.33483	X	X			X		X		
5	5.0	2.0	2.4	0.33419	X	X			X		X	X	
5	4.8	1.8	2.8	0.33460	X	X					X	X	X
6	5.2	1.6	4.1	0.33487	X	X			X		X	X	X
6	5.1	1.5	4.3	0.33511	X	X			X	X	X	X	
7	5.2	1.0	6.0	0.33586	X	X			X	X	X	X	X
7	5.2	1.0	6.1	0.33591	X	X		X	X		X	X	X
8	5.3	0.4	8.0	0.33691	X	X		X	X	X	X	X	X
8	5.2	0.4	8.0	0.33694	X	X	X		X	X	X	X	X
9	5.3	0.0	10.0	0.3800	X	X	X	X	X	X	X	X	X

Table 5: Best subsets regression: log (C) versus all factors. R^2 , adjusted R^2_{adj} , Mallow's C_p (MCP) and standard error (SE) in each subset is presented. V denotes the number of factors included within each model.

those subset models that fall near the line $C_p = p$ in a plot of C_p against p for the collection of all subset models under consideration. Under this view, only the combination of all factors except factor G (Ground Type, Table 3) constitutes an explanatory subset for minimising total bias.

Additionally to the analysis presented so-far, one-way ANOVA was applied to single-factor data from the whole data set. Analysing factor Level-L in its full depth, an f value of 2.551 was calculated whereas the critical f value at the 95% confidence interval-CI is 2.03. These values imply that with P=0.014 results from the various different levels collected did not differ significantly. However when the full dwelling-level data were reorganised in 3-levels (ground floor, first floor and upper floors) f value was found equal to 7.156 while the corresponding critical value at the 95% CI is 3.01 with corresponding value P=0.000877. This finding is significant since it implies that with P<0.001 lower level dwellings present higher indoor radon concentrations. Further evidence was provided by reorganising dwelling-level data in 2 levels,

namely ground floor and higher floor dwellings. Applying t-test to the average concentrations it was calculated that at $p<0.001$ ground floor dwellings presented higher radon concentrations. Analysing factor "Wall Contact-Con" with one-way ANOVA, an f value of 0.893 was calculated whereas the critical f value at the 95% CI is 2.624. Namely, at $p<0.001$ different Wall Contact - factor level dwellings did not present differences. Similar was the outcomes of the one-way ANOVA for factor "Floor's material-F". Non-significant variations were addressed, since calculated f value was 1.298 and the critical value at 95% CI, 2.119. On the contrary, the one-way analysis of factor "Wall's material-W", provided an f value of 4.314 with a critical value at 95% CI of 2.624 and an associated *P* value of 0.0051. The latter finding was associated with a tendency of higher concentrations of rock dwelling.

Factor	p value
Area-A	0.131
Building Type-BTy	0.162
Floor's material-F	0.185
Level-L	0.208

Table 6: General MANOVA for log(C). Cut-off Limit for significance level < 30%

Step	1	2	3
Constant	1.438	1.467	1.487
Wall Contact-Con	**-0.070**	**-0.072**	**-0.073**
T-Value	-1.34	-1.38	-1.40
P-Value	0.182	0.168	0.162
Level-L	**-0.048**	**-0.055**	
T-Value	-1.35	-1.52	
P-Value	0.178	0.131	
Wall's material-W	**-0.040**		
T-Value	-1.09		
P-Value	0.276		
S	0.337	0.336	0.336
R^2	1.09	2.20	2.92
R^2_{adj}	0.48	0.99	1.11
MCP	-1.2	-1.0	-0.2

Table 7: Stepwise Regression: log (c) versus all factors. Value of Alpha-to-Enter was 0,5 and of Alpha-to-Remove 0,5. P-values represent calculated error probabilities and T-values, the corresponding values of t-student's test for the comparison of a P-value with a . The constants of the linear fit at each step are shown in first row. Bold values represent the corresponding slopes of each factor at each step. MCP is the Mallow's C_p .

Table 6 presents the results of the general MANOVA method. These results support further findings of the one-way ANOVA. Non-significant statistical interactions between any combinations of factors were detected by General MANOVA

Table 7 presents the results of stepwise regression of log (*C*)versus all factors. Through stepwise regression, a linear model was sought containing only those variables which were significant in modelling log (*C*). The qualitative factor levels of Table 3 were employed in their original values so as to be transformed to quantitative variables. It should be stressed, that stepwise regression is particularly useful when there are many possible explanatory (independent) variables. Some of these variables may be highly correlated with each other and therefore may explain the same variation in the response and not be independently predictive. Some may also not influence the response in any meaningful way. For this reason, employing high alpha values for the error probability in Table 7, only three factors were finally selected, and these after the third step. It is noted that probability values, p , and alpha values, a , are related as $p= 1- a$ and thus the results of Table 7 correspond to the 50% significance level either for accepting entering of a certain factor or its removal. According to Table 7 the main factors found to influence indoor radon concentrations were "Wall's material- W", "Level- L" and "Wall Contact- Con". In specific, factor contact exhibited p -value of 0.162, factor level, 0.131 and factor wall 0.276 (error probability 50%). This implies that at a significance level <17% level and contact affect indoor radon concentration, while at the 30% significance level, all three factors affect. These results, however, provide vague evidence on the null hypothesis, namely that the above factors actually affect. This is also indicated by the small value of MCP in reference to an accepted well value of 3 for two factors and 4 for 3 factors. Moreover, since R^2 exhibited maximum value of 2.92, only a small percent of the total variance (<3%) can be described by a linear model of the three factors of Table 7.

According to data presented so-far, no single factor or linear subset of factors, could describe sufficiently the variance of the analysed radon concentration data. This implies that a multivariate set of factors could be probably more adequate. Table 8 presents the unrotated four principal factor loadings together with the corresponding communalities according to principal component analysis, applied however to radon concentrations. It is very interesting that, although factor 1 is loaded to five factors (bold numbers), the remaining three are only loaded to one single independent factor each. More specifically, 16.2% of the total variance may be described by factor 2 loaded only to the construction year. 12.3% of the total variance could be attributable to factor 3 loaded mainly to the existence of basement and 10.8% to the existence

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Communality
C	0.178	0.104	0.444	-0.415	1.000
Area-A	-0.463	-0.156	-0.480	0.061	1.000
Level-L	**-0.588**	-0.415	-0.157	-0.44	1.000
Ground Type-G	0.242	0.410	-0.498	0.026	1.000
Basement-Bas	0.291	0.405	**-0.646**	-0.111	1.000
Building Type-BTy	**-0.647**	-0.416	-0.236	-0.204	1.000
Year-Y	**-0.527**	**0.598**	0.088	-0.119	1.000
Wall Contact-Con	-0.123	-0.386	0.000	**0.750**	1.000
Wall's material-W	**0.690**	-0.428	-0.213	-0.127	1.000
Floor's material-F	**0.548**	-0.472	-0.080	-0.266	1.000
Variance	2.2225	1.6241	1.2337	1.0891	10.0000
% Var	0.222	0.162	0.123	0.109	1.000

Table 8: Unrotated factor loadings and communalities for 4 principal factors. The factor correlations exceeding the cut-off limit of 0.5 are marked in bold.

of contact (factor 4). A very important finding of Table 8, however is that since the loadings of factor 1 to "Level- L", "Building Type" and "Construction Year- Y" are negative in respect to the one of *C*, it is rational to accept that concentrations would increase as Level, Building Type and Year are decreased. According to Table 3, this implies that ground floor dwellings tend to present higher radon concentrations. This is rational since the lower the floor, the higher is the contribution of soil's exhalation in indoor radon. Also detached houses tend to present higher concentrations, since other types offer pathways for radon's interchange between dwellings in contact. Significant is also that aged dwellings, especially those of the previous century, presented higher radon concentrations. The latter finding is also reinforced by the positive loading of the "Wall's material-W", especially due to its rather high loading. Since higher wall values correspond to rock materials, it can be supported that higher concentrations are addressed in dwellings of the beginning of the twentieth century made of rocks. To some degree these results were supported by Table 7, since the dwelling's "Level-L" and building "Wall's material-W" were considered to be more significant compared to other factors.

Conclusion

Statistical analysis of data revealed that approximately 0.1% of the dwellings exhibited outlier radon concentrations. Noteworthy statistical correlations were detected between indoor radon concentration and the factors: "Level-L" and "Wall's material-W". Results of current work strengthened these considerations and provided weak evidence for the correlation of factors like "Building Type-Bty", "Construction Year-Y" and "Floor's material-F" with radon concentrations. Minor association was detected with the factor "Wall Contact-Con". Significant differences were detected in results produced by some of the applied statistical methods.

Acknowledgement

This work was co-financed by Greece and the European Union under the European Social Fund NSRF 2007-2013 (Thales). Managing Authority: Greek Ministry of Education & Religious Affairs, Culture & Sports.

Financial Support-Funding

This work has been co-financed by Greece and the European Union, under the European Social Fund NSRF 2007-2013 (Thales). Managing Authority: Greek Ministry of Education and Religious Affairs, Culture and Sports.

References

1. Kucukomeroglu B, Maksutoglu F, Damla N, Cevik U, Celebi N (2012) A study of environmental radioactivity measurements in the Samsun province, Turkey. Radiat Prot Dosimetry 152: 369-375.
2. Ngachin M, Garavaglia M, Giovani C, Kwato Njock MG, Nourreddine A (2008) Radioactivity level and soil radon measurement of a volcanic area in Cameroon. J Environ Radioact 99: 1056-1060.
3. Clouvas A, Xanthos S, Takoudis G (2011) Indoor radon levels in Greek schools. J Environ Radioact 102: 881-885.
4. Baan RA, Grosse Y (2004) Man-made mineral (vitreous) fibres: evaluations of cancer hazards by the IARC Monographs Programme. Mutat Res 553: 43-58.
5. ICRP 65 (International Commission on Radiological Protection) 1994 Protection against radon-222 at home and at work. Oxford: Pergamon Press
6. UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation) 2006 Annex E: sources-to-effects assessment for radon in homes and workplaces. New York: United Nations.
7. WHO 2009 Handbook on indoor radon: a public health perspective / edited by Hajo Zeeb and Ferid Shannoun. ISBN 978 92 4 154767 3.
8. Kozak K, Mazur J, Kozowska B, Karpinska M, Przylibski TA, et al. (2011) Correction factors for determination of annual average radon concentration in dwellings of Poland resulting from seasonal variability of indoor radon. Applied Radiation and Isotopes 69: 1459–65
9. Tondeur F, Rodenas J, Querol A, Ortiz J, Juste B (2011) Indoor radon measurements in the city of Valencia. Appl Radiat Isot 69: 1131-1133.
10. Kurnaz A, Kucukomeroglu B, Cevik U, Celebi N (2011) Radon level and indoor gamma doses in dwellings of Trabzon, Turkey. Appl Radiat Isot 69: 1554-1559.
11. Rafique M, Rahman S U, Mahmood T, Rahman S, Matiullah (2011) Assessment of Seasonal Variation of Indoor Radon Level in Dwellings of Some Districts of Azad Kashmir, Pakistan Indoor Built Environ 20 354–61.
12. Kim Y, Chang BU, Park HM, Kim CK, Tokonami S (2011) National radon survey in Korea. Radiat Prot Dosimetry 146: 6-10.
13. Ramola RC (2011) Survey of radon and thoron in homes of Indian Himalaya. Radiat Prot Dosimetry 146: 11-13.
14. Mehra R, Badhan K, Kansal S, Sonkawade RG (2011) Assessment of seasonal indoor radon concentration in dwellings of Western Haryana. Radiation Measurements 46: 1803-1806.
15. Stojanovska Z, Januseski J, Bossew P, Zunic Z S, Tollefsen T, et al. (2011) Seasonal indoor radon concentration in FYR of Macedonia. Radiation Measurements 46: 602-610.
16. Miles JC, Howarth CB, Hunter N (2012) Seasonal variation of radon concentrations in UK homes. J Radiol Prot 32: 275-287.
17. Szeiler G, Somlai J, Ishikawa T, Omori Y, Mishra R, et al. (2012) Preliminary results from an indoor radon thoron survey in Hungary. Radiat Prot Dosimetry 152: 243-246.
18. Valmari T, Arvela H, Reisbacka H (2012) Radon in Finnish apartment buildings. Radiat Prot Dosimetry 152: 146-149.
19. Cucos Dinu A, Cosma C, Dicu T, Begy R, Moldovan M, et al. (2012) Thorough investigations on indoor radon in Baita radon-prone area (Romania). Sci Total Environ 431: 78-83.
20. Proukakis C, Molfetas M, Ntalles K, Georgiou E, Serefoglou A (1988) Indoor radon measurements in Athens, Greece. In: British Nuclear Energy Society Conference on health effects of low dose ionizing radiation-recent advances and their implications. BNSE, London, 177–178.
21. Georgiou E, Ntalles K, Molfetas M, Athanassiadis A, Proukakis C (1988) Radon measurements in Greece. In Radiation Protection Practice. New York, Pergamon Press, IRPA 7, 1: 125–126.
22. Papastefanou C, Stoulos S, Manolopoulou M, Ioannidou A, Charalambous S (1994) Indoor radon concentrations in Greek apartment dwellings. Health Phys 66: 270-273.
23. Ioannides KG, Stamoulis KC, Papachristodoulou CA (2000) A survey of 222Rn concentrations in dwellings of the town of Metsovo in north-western Greece. Health Phys 79: 697-702.
24. Clouvas A, Xanthos S, Antonopoulos-Domis M (2007) Pilot study of indoor radon in Greek workplaces. Radiat Prot Dosimetry 124: 68-74.
25. Clouvas A, Xanthos S, Takoudis G (2011) Indoor radon levels in Greek schools. J Environ Radioact 102: 881-885.
26. Nikolopoulos D, Louizi A, Koukouliou V, Serefoglou A, Georgiou E, et al. (2002) Radon survey in Greece--risk assesment. J Environ Radioact 63: 173-186.
27. Baixeras C, Font L, Robles B, Gutierrez J (1996) Indoor radon survey in the most populated areas in Spain Environ. Int 22: S671–6
28. Gallelli G, Panatto D, Lai P, Orlando P, Risso D (1998) Relevance of main factors affecting radon concentration in multi-storey buildings in Liguria (Northern Italy). J Environ Radioact 39: 117–128.
29. Bochicchioa F, Campos-Venutia G, Piermatteib S, Nuccetellia C, Risicaa S, et al. (2005) Annual average and seasonal variations of residential radon concentration for all the Italian Regions. Radiat Meas 40: 686–694.
30. Bossew P, Lettner H (2007) Investigations on indoor radon in Austria, Part 1: Seasonality of indoor radon concentration. J Environ Radioact 98: 329-345.
31. Denman AR, Groves-Kirkby NP, Groves-Kirkby CJ, Crockett RGM, Phillips

PS, et al. (2007) Health implications of radon distribution in living rooms and bedrooms in U.K. dwellings - a case study in Northamptonshire. Environment International 33: 999-1011.

32. Cosma C, Cucos-Dinu A, Papp B, Begy R, Sainz C (2013) Soil and building material as main sources of indoor radon in Baitei radon prone area (Romania). J Environ Radioact 116: 174-179.

33. Bavarnegin E, Fathabadi N, Vahabi Moghaddam M, Vasheghani Farahani M, Moradi M, et al. (2013) Radon exhalation rate and natural radionuclide content in building materials of high background areas of Ramsar, Iran. J Environ Radioact 117: 36-40.

34. Kenney JF, Keeping ES (1962) Linear Regression and Correlation Ch. 15 in Mathematics of Statistics, Pt. , 3rd ed. Princeton, NJ: Van Nostrand 252-85

35. Young D (2013) Regression Methods STAT 501 Pennsylvania State University

36. Kéry M (2010) Introduction to WinBUGS for Ecologists. 9: 115-27

37. Manousakas M, Fouskas A, Papaefthymiou H, Koukouliou V, Siavalas G, et al. (2010) Indoor radon measurements in a Greek city located in the vicinity of lignite-fired power plants. Radiation Measurements 45: 1060-1067.

38. Papachristodoulou CA, Patiris DL, Ioannides KG (2010) Exposure to indoor radon and natural gamma radiation in public workplaces in north-western Greece. Radiation Measurements 45: 865-8671

39. Trevisi R, Leonardi F, Simeoni C, Tonnarini S, Veschetti M (2012) Indoor radon levels in schools of South-East Italy. J Environ Radioact 112: 160-164.

40. Ju Yong-Jin, Ryu Young-Hwan, Dong Kyung-Rae, Cho Jae-Hwan, Lee Hae-Kag , et al. (2012) Study on measurement and quantitative analysis of Radon-222 emitted from construction materials. Annals of Nuclear Energy 49: 88–95.

41. Mathew T (1989) MANOVA in the Multivariate Components of Variance Model. J Multivar Anal 29: 30-38.

42. Stahle L, Wold S (1990) Multivariate Analysis of Variance (MANOVA). Chemometrics and Intelligent Laboratory Systems 9: 127-141.

43. Montgomery D, Runger G (1994) Applied Statistics and Probability for Engineers Wiley

44. Hauri DD, Huss A, Zimmermann F, Kuehni CE, Roosli M (2012) A prediction model for assessing residential radon concentration in Switzerland. J Environ Radioact 112: 83-89.

45. Appleton JD, Miles JCH, Young M (2011) Comparison of Northern Ireland radon maps based on indoor radon measurements and geology with maps derived by predictive modelling of airborne radiometric and ground permeability data. Science of the Total Environment 409: 1572–1583.

46. Neuberger JS, Mahnken JD, Mayo MS, Field RW (2006) Risk factors for lung cancer in Iowa women: implications for prevention. Cancer Detect Prev 30: 158-167.

47. Naes T, Isaksson T, Fearn T, Davies T (2002) A User-Friendly Guide to Multivariate Calibration and Classification NIR Publications Chichester

48. Sanguansat P (2012) Principal Component Analysis – Multidisciplinary Applications, Rijeka, Croatia. In Tech, ISBN 978-953-51-0129-1.

49. Toranzos GA, McFeters GA (1997) Remove from marked Records. Detection of indicator microorganisms in environmental freshwaters and drinking waters. In Manual of environmental microbiology 184-194.

50. McWilliams RR, Bamlet WR, de Andrade M, Rider DN, Cunningham JM, et al. (2009) Nucleotide excision repair pathway polymorphisms and pancreatic cancer risk: evidence for role of MMS19L. Cancer Epidemiol Biomarkers Prev 18: 1295-1302.

51. Chang H, Kim D (2010) A quality function deployment framework for the service quality of health information websites. Healthc Inform Res 16: 6-14.

52. Singh A, Kumar A, Kumar A (2013) Determinants of neonatal mortality in rural India, 2007-2008. PeerJ 1: e75.

53. Tanaskovic I, Golobocanin D and Miljevic N (2012) Multivariate statistical analysis of hydrochemical and radiological data of Serbian spa waters Journal of Geochemical Exploration 112: 226–34

54. Johnson R and Wichern D (2002) Applied Multivariate Statistical Analysis. Prentice Hall, New Jersey pp 426-76

55. Skeppström K (2005) Radon in groundwater - influencing factors and prediction methodology for a Swedish environment Land and Water Resources Engineering ISBN 91-7178-208-7 KTH

56. Nikolopoulos D, Louizi A, Petropoulos N, Simopoulos S and Proukakis C (1999) Experimental study of the response of cup-type radon dosemeters Radiation Protection Dosimetry 83: 263–6