



Statistical Study for the Energy Resolution of the Cesium Source Using the NaI(Tl) Detector

Inaam H. Khadim^{1*} and Khalid Hussain Hatif²

^{1,2} Collage of Science, University of Babylon , Iraq

*Email: stud.inaam.hadi@uobabylon.edu.iq

*Corresponding author email: Email : stud.inaam.hadi@uobabylon.edu.iq

دراسة إحصائية لقابلية تحليل الطاقة لمصدر السيزيوم باستخدام كاشف NaI (Tl)

انعام هاني كاظم^{1*}, خالد حسين هاتيف²

¹ كلية العلوم، جامعة بابل، Email: stud.inaam.hadi@uobabylon.edu.iq، الحلة، العراق

Received:	5/2/2021	Accepted:	12/2/2022	Published:	1/3/2022
-----------	----------	-----------	-----------	------------	----------

ABSTRACT

To investigate the statistical nature of gamma-ray decay, the energy resolution of the ¹³⁷Cs source spectrum was investigated using the statistical programming language R, as well as the moderation and homogeneity test for the data, as well as the Pearson correlation coefficient test and the linear regression test. It was observed that the investigated random variables had no strong statistical link, and that the random mean of the samples differs slightly from the expected mean for same variables using same statistical distribution.

Materials and Methods:

In this study, a scintillation detector of iodide sodium activated with thallium NaI(Tl) measuring (200) samples of ¹³⁷Cs spectrum at (200 sec) time for each spectrum and a distance is(7cm) between the radioactive source and the detector . The R programming language was used in calculating the statistical applications represented by the mean, median, variance and standard deviation, also the normality tests and homogeneity of variances tests were calculated by the R programming language, it had been found that all the models studied were homogeneous and that most of the models were not follow to a normal distribution.

Conclusions:

In this study, statistical distributions and statistical software were used, and it was revealed to the following:

- 1-The Kolmogorov-Smirnov test revealed that the data in the sampling random analysis are distributed in a regular manner.
- 2-The R.E. of the spectrum data was found to follow a normal distribution in a Shapiro test, depending on the p. value.
- 3-In the Q.Q. Normal Test, the points in R.E are applied in a straight line, indicating that the data is homogeneous.
- 4-Using the linear regression test, it was determined that the analyzed random variables have no statistically significant correlation.
- 5-It was discovered that the analyzed samples (R.E) follow a normal distribution using the T-Test.



6-The third group in Pearson correlation was less than zero, while the rest of the groups were near zero, with the first and second groups having negative values, indicating that the investigated variables have a weak inverse link. The third group had a positive correlation value, indicating that the ^{137}Cs source's R.E. had increased.

Key words:

^{137}Cs source, R programming language, linear regression test, gamma-ray decay.

الخلاصة

في هذه الدراسة، تم استخدام قوانين الإحصاء وبرامج التوزيع الإحصائي باستخدام لغة البرمجة R لمحاولة فهم الطبيعة الإحصائية العشوائية لانحلال أشعة كاما وتفاعلها مع المادة من خلال دراسة قابلية تحليل الطاقة لطيف مصدر ^{137}Cs بالإضافة إلى استخدام اختبار الاعتدال واختبار تجانس التباينات وكذلك اختبار معامل ارتباط بيرسون واختبار الانحدار الخطي. وقد وجد أنه لا توجد علاقة إحصائية قوية بين المتغيرات العشوائية المدروسة، كما وجد أن هناك اختلاف بسيط بين المتوسط العشوائي لقراءات العينات المقاسة والقراءات المتنبئ بها من قبل البرنامج لنفس العينات باستخدام نفس التوزيع الإحصائي.

طرق العمل:

تمت هذه الدراسة باستخدام كاشف يوديد الصوديوم المنشط بالثاليوم (TI) NaI بحجم بلورة (3 "x" 3) حيث جمعت 200. طيف لمصدر السيزيوم ووقت التجميع لكل طيف (200 ثانية) والمسافة (7 سم) بين المصدر المشع والكاشف وتم استخدام لغة البرمجة R في حساب التطبيقات الإحصائية المتمثلة بالمتوسط والوسيط والتباين والانحراف المعياري، كما تم حساب اختبارات المعيارية واختبارات تجانس التباينات بلغة البرمجة R، وقد وجد أن جميع النماذج المدروسة كانت متجانسة وأن معظم النماذج لم تكن تتبع التوزيع الطبيعي.

الاستنتاجات:

- في هذه الدراسة تم استخدام التوزيعات الإحصائية والبرمجيات الإحصائية، وتبين لنا ما يلي:
- 1- في اختبار Kolmogorov-Smirnov وجد أن البيانات المدروسة تخضع للتوزيع الطبيعي.
 - 2- في اختبار Shapiro، وجد أن R.E لبيانات الطيف كان يتبع التوزيع الطبيعي اعتمادًا على p.value.
 - 3- في اختبار Q-Q، نلاحظ أن النقاط في R.E يتم تطبيقها على طول خط مستقيم، وبالتالي فإن البيانات متجانسة.
 - 4- باستخدام اختبار الانحدار الخطي وجدنا أنه لا توجد علاقة إحصائية قوية بين المتغيرات العشوائية المدروسة.
 - 5- باستخدام اختبار T-Test وجد أن العينات المدروسة تخضع للتوزيع الطبيعي.
 - 6- وجد باستخدام معامل ارتباط بيرسون أن هناك علاقة عكسية ضعيفة بين المتغيرات المدروسة.

الكلمات المفتاحية:

أشعة كاما، الكاشف الوميضي، التوزيع الإحصائي، لغة البرمجة R، معامل ارتباط بيرسون.

INTRODUCTION

There are some phenomena in nuclear physics that do not have a direct explanation and are not subject to clear laws, so they are based on probability and statistics theory [1]. These phenomena include radioactive decay [2], decay of compound nuclei [3], nuclear fission products [4], and nuclear reactions [3]. There are many studies on this topic from a statistical point of view [1].

Energy resolution (R.E.) is a measure of the ability of the counter to differentiate between two closely energies [2]. The quantum line width at its mid-peak



height (FWHM) (Full Width at Half Maximum), to measure the energy separation of the NaI(Tl) detector, was use [4].

$$\text{Energy Resolution (R)} = \text{FWHM} / (\text{P.P}) \times \% \quad (1)$$

where:

FWHM: the width at half of the peak level.

p.p: the centroid of photopeak.

A scintillation detector was used to explore the statistics of independent random events in physical measurements of gamma rays. And obtaining our experimental results, then using statistical and software distributions on experimental data as an attempt to understand and explain the random nature and monitor these nuclear phenomena that are subject to any law and then discuss this result. The aim of the research is applying statistical models and global codes to try to interpretation the random behavior of nuclear phenomena because we believe that there is no randomness, but rather the existence of accurate laws. so, the subject of statistics and probability cannot reach interpretations. and that is through practical results by us and apply the latest codes in this topic, such the language R programming statistical.

Theoretical Part

Statistical computing represented the application of computer science to statistics [5]. It's used in a variety of fields, sciences, and statistics, including data analysis, graphics, simulations, etc. In addition to mathematically intensive statistical methods including reduction methods, Markov chain Monte Carlo methods and linear regression and others. Among the programs specialized in statistical computing are (MATLAB, R, S, Excel, SPSS) [4].

In this study, used the latest available version of the R system which is a 4.0.3 version. One of the most well-known statistical programs is the R-R program. The R programming language was created by Ross Ihaka and Robert Gentleman of the University of Auckland in New Zealand and is used in this application. [5]. R Studio is a user interface for the R statistical software that includes a number of useful windows and features to help you organize your programming session.as shown in the figure (1).



Normality Tests and Homogeneity of Variances Tests

When testing any hypothesis, it must be formulated as two hypotheses, hypothesis (Hypothesis Null) H_0 , and the second hypothesis is (Hypothesis Alternative) H_1 , and the goal of testing the hypothesis is to study the possibility of rejecting the primary hypothesis at a specific level of significance. The alternative and determining the level of importance, usually 0.05 [6]. so, finds calculating the appropriate test and calculating the significance p.value, then comes a stage take a statistical decision by comparing the test's significance with the level of importance, if it is valuable If the p.value is less than (0.05), the initial hypothesis is rejected; however, if the p.value is larger than the threshold of statistical significance, the first hypothesis cannot be discarded. [7].

Hypotheses Test is employed, as Eq. [6].

opposite

$$H_1: X \neq N(\mu, \sigma^2) \quad (2)$$

Q-Q Plot

Page | 36



around a normal straight line, the data is normal, and whenever the data moves away from the normal distribution line, this indicates a lack of commitment the data is in a normal distribution [7].

Bartlett's Test

It's utilized when the data follows a normal distribution and divided into several groups according to a specific factor and to test the hypothesis, as Eq. [7].

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_r^2$$

opposite

$$H_1: \sigma_i^2 \neq \sigma_j^2 \text{ , for some } i, j \quad (3)$$

T-Test Independent Samples

T -Test is used for the two independent samples to test the hypothesis, as Eq. [8]

$$H_0: \bar{x} - \bar{y} = m$$

opposite

$$H_1: \bar{x} - \bar{y} \neq m \quad (4)$$

Which studies the significance of the difference between $\bar{x} - \bar{y}$ equals a fixed value such as m.

Coefficient Correlation Pearson

The Pearson Correlation Coefficient is used to study the presence of a relationship between two quantitative variables X and Y and the intensity of these the relationship, whenever its absolute value approaches One, this indicates that the relationship is stronger, but when its absolute value approaches the half, it is the relationship is medium, and when its value approaches zero, the relationship is weak, and the positive sign is the correlation coefficient indicates that the relationship is direct, while the negative sign indicates that the relationship inversely, after computing the coefficient of correlation (R) the hypothesis is tested, as Eq [9]:

$$H_0: R = 0$$

opposite

$$H_1: R \neq 0 \quad (5)$$

Simple Linear Regression

Simple linear regression is the study of the effect of the variable X, (Variable Independent) on the Y (Variable Dependent) [10]. The linear regression model is given as Eq. [10]:

$$Y = \beta_0 + \beta_1 X + \epsilon \quad , \epsilon; \epsilon \sim N(0, \sigma^2) \quad (6)$$

where ϵ is residual or error.

The regression analysis aims to find estimations for each of β_0, β_1 , This makes the sum of the squares of sediments as small as possible apply simple linear regression using r by the instruction $\text{lm}(y \sim x)$ [11].

Multiple Linear Regression

Multiple linear regression is the study of the effect of several variables $\beta_1, \beta_2, \dots, \beta_P$ called (Variables Independent) on a y variable called the Dependent Variable, the multiple linear regression model is given as, as Eq. [12].

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_P X_P + \epsilon; \epsilon \sim N(0, \sigma^2) \quad (7)$$

Materials and Methods:

The NaI(Tl) detector, which was provided by a corporation (Spectrum Techniques LLC), used an electronic counting and analysis system, as illustrated in figure (2). . It was chosen due to the materials' high gamma ray penetration strength. The UCS-30 computer program was used to perform the nuclear measurements and analysis.

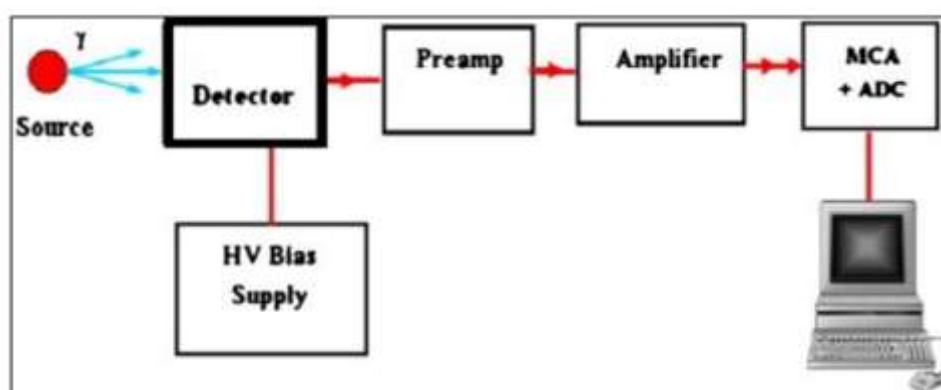
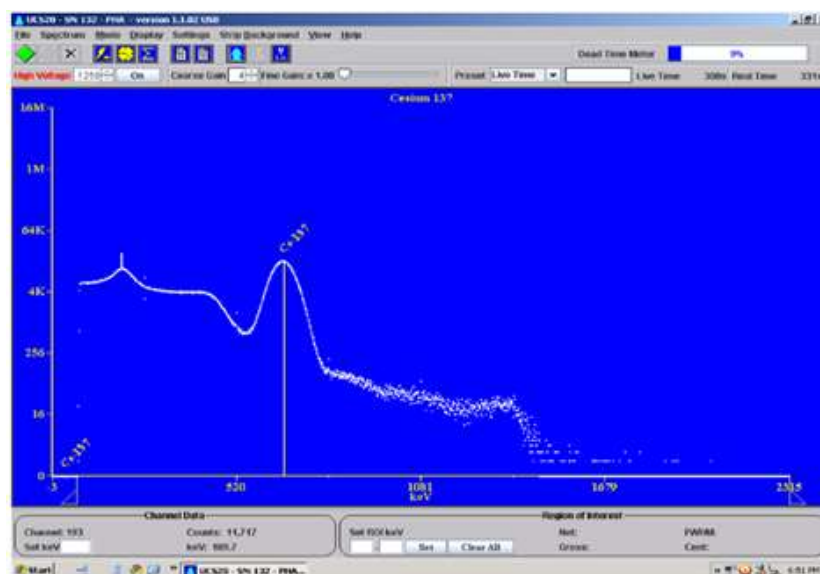


Figure (2) Nuclear Counting Detector .

For half-life and decay studies using the UCS30 ADVANCED Spectrometer System comprises a 4096 channel MCA with integrated preamplifier, high voltage (0-2048V), higher and lower-level discriminators, and multichannel scaling. [1].

The scintillation detector of iodide sodium was activated with thallium NaI(Tl) at size in this investigation (3x 3)inch. As seen in figure (3), the spectrum created by the emission of these photons has a single photopeak. A total of 200 samples were collected. The ^{137}Cs spectrum has a collecting time is 200 seconds, and the distance between the radioactive source and the detector is 7centimeters.



Results and Discussion

The spectrum was measured for ^{137}Cs source after and before background in terms of Energy Resolution (R.E.) as table (1) .

Table (1) summary ^{137}Cs source spectrum

summary		R.E.	
		Before	After
Population	Min	0.04777	0.06613
	Max	0.07772	0.07446
	1st Qu	0.06943	0.06926
	Median	0.07049	0.07043
	3rd. Qu	0.07178	0.07171
	IQR	0.00235	0.00245
	O- up	0.081245	0.078135
	O-down	0.044245	0.069805
	Mean	0.07048518	0.07043431
	Var	6.062001e-06	2.888173e-06
	σ	0.002462113	0.001699463
Sample	(EX)	0.07050699	0.07042285
	Var	6.075703e-06	2.903885e-06

	sd	0.00246489	0.00170408
	Median	0.07085	0.07051
	IQR	0.00233	0.00246
	O- up	0.081215	0.077955
	O-down	0.044275	0.062635
Sample standard Bar	X.bar	0.07048523	0.07043308
	Var	3.442291e-10	1.519468e-10
	sd	1.855341e-05	1.232667e-05
	Median	0.07048	0.07043
	IQR	1e-05	3e-05
	O-up	0.074095	0.074095
	O-down	0.066955	0.066955

From table (1). shows a summary for R.E to ^{137}Cs source as figure (4) that show the behavior of the sample.

where:

IQR : is the distance between the third and first quartiles .

μ : the average of the population's values

var : a sample's standard deviation

σ : the population's variation

x.bar : is standard mean

sd : The standard deviation is a measurement

o-up : large outliers

o-down : small outliers

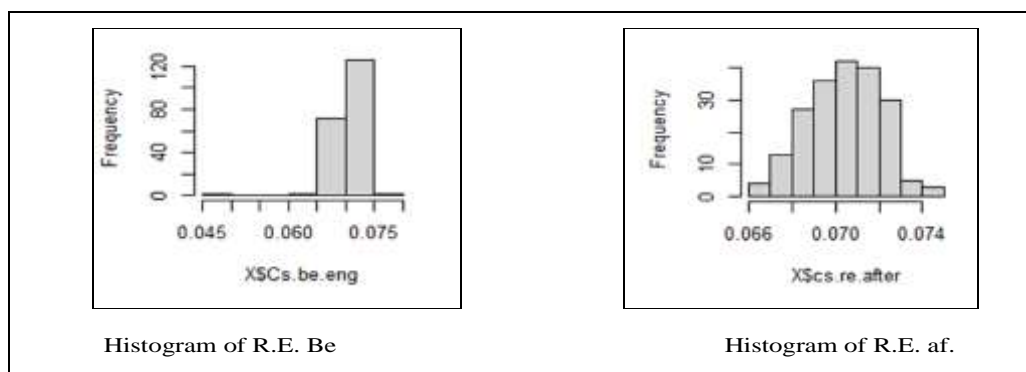


Figure (4) The General Behavior for Energy Resolution.



Where; R.E.Be & R.E.af are energy resolution before and after background respectively. The histogram shows the general behavior of the spectrum, where the x-axis represents the R.E, while the number of repetitions of the data is on the y-axis, we notice R.E.Be difference from the R.E.af. Thus, the histogram can be used to estimate the data's distribution and dispersion around the mean., The number of observations that fall inside the mean, which has been divided into equal periods, is indicated by the height of each box.

Compare the Average for Two Sets of Data Using R Programming Language

The Normality Test, which includes the Kolmogorov-Smirnov Test, Shapiro.Test, T-Test, and Bartlett Test for data samples, is used to determine the homogeneity of the variance, as table (2).

Table (2) Test of Normality and Homogeneity of Variances.

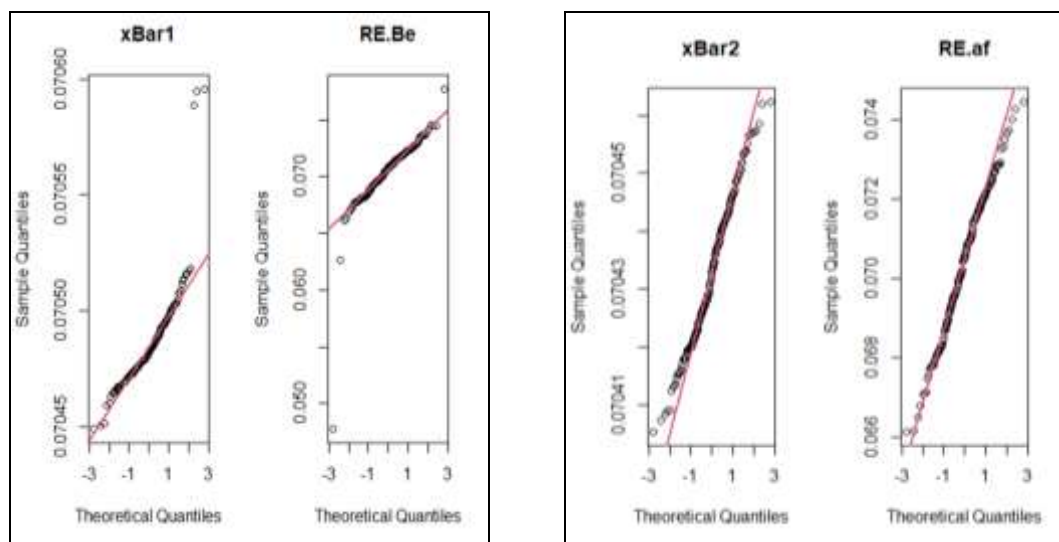
Type of Test	Kolmogorov-Smirnov Test	Shapiro.Test	Bartlett. Test	T-Test	Wilcoxon.Test
R.E.Be	0.038	$2.2e^{-16}$	0.755	0.991	0.317
R.E.af	0.417	0.189	0.592	0.998	0.801

From table (2) that show the normality test determines if the data follows a normal distribution or not. Because the p.value is bigger than the significance level of 0.05, we accept to impose nothingness Null hypothesis (H_0) and if it is below the moral level (0.05) and therefore we refuse to the decision is that the two groups are left behind in relation to the average, as shown in (Eq.8). Therefore, it has been found that the total area samples are not follow to a normal distribution because the p.value is less than 0.05, while the photopeak area samples were less than 0.05 before the background, while samples after the background were greater than 0.05. so, is the ratio valley sample like for the energy resolution, i.e., the p-value was larger than 0.05 before and after the background radiation, thus it is following a normal distribution.

According to Shapiro's test, R.E.Be does not follow a normal distribution dependent on the p. value, whereas R.E.af follows a normal distribution.

The p.value in the Bartlett test is greater than 0.05, implying that the variance of the two groups is roughly equivalent at the level of significance 0.05, allowing us to conclude that the data in all samples analyzed is homogenous.

In T-Test that used to compare mean by two groups, where the alternative hypothesis is the actual difference of the samples is not equal to zero. Therefore, found through the p.value greater than 0.05 this mean we accept the primary hypothesis, so the mean of the samples does not differ substantially, meaning that the samples are statistically regular, as figures (5). which the comparative of sampling means by t-test. where: (\bar{x}_1 , \bar{x}_2) are assumed samples for data sampling.



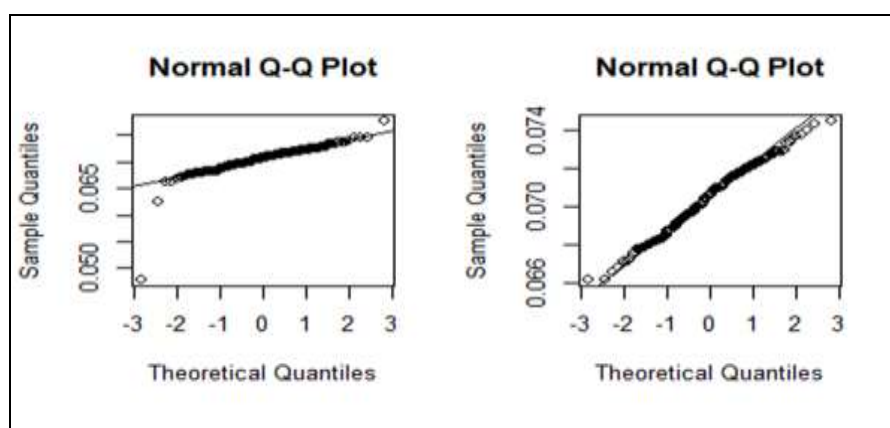
a

b

Figure (5) The Comparative of Energy Resolution (a- Before and b- After) Sampling Means By T-Test.

Q.Q. Normal Test

For make sure the imposition is whether the data given follows the normal distribution. We will use Q-Q Plot, as shown in figures (6).



a

b

Figure (6): Q-Q Norm for Data Samples (R.E). (a- Before and b- After) Background

It have been said figures (5) that the data follows the natural distribution from its notice the points are applied along a straight line.

Pearson correlation

In this study, Pearson correlation coefficient was used to study the existence of a relationship between the variables of the studied samples using the R



program. The studied data were divided into two groups before and after the background, as shown in a table (3).

Table (3) Pearson correlation.

Data	p-value	Correlation	correlation	confidence interval	
Cs.before.R.E and Cs.R.E.after	$2.2e^{-16}$	0.5959189	Positive	0.4983834	0.6785579
sample and Cs.before.R.E	0.2713	-0.07815767	negative	-0.21457194	0.06124749
sample and Cs.R. E.after	0.2089	-0.08924087	negative	-0.22519368	0.05012071

It was noticed from the table (3) that the correlation coefficient was negative and p.value greater than 0.05 for the group (samples and R.E.), unlike the group R.E. before and after background whose correlation coefficient positive and p.value less than 0.05 which indicates there is no essential relationship between the two groups.

Linear Regression Test

The concept of correlation and regression relates to the study of linear or non-linear relationships between variables, and regression models reconcile a causal relationship between these variables and are dealt with by calculating a numerical scale that reflects the degree and type of relationship between the two variables as shown in the table (4).

Table (4) Linear Regression Test.

Data	p-value	Std-Error	Estimated	Min.	Max.	Median	R.squared
Data.en_\$sample ~ Cs.be.eng	0.2713	1665.5	0.5959189	-100.500	98.516	2.622	0.001089
Data.en_\$sample ~ cs.re.after	0.2089	2410.7	-3039.3	-103.127	99.336	2.606	0.002954
Data.en.`Cs be.eng` ~ Xcs.re.after	$2e^{-16}$	0.082681	-0.089240	-0.02411	0.0071061	0.0000161	0.3519

The simple linear regression of the studied samples can be seen in table (4), which contains the lowest value, highest value, mean error, coefficient of determination value, and residual value, which represents the difference between the actual value and the predicted value; the closer its value to zero, the better the model's quality.

The p.value, which shows the level of significance, is also included in the table; the findings revealed that most of the groups had p. values larger than 0.05, while group three's p. value was less than 0.05, meaning that the first group is more relevant than the rest. So, it was used a multiple linear regression test to see which groupings influence the other variables, as shown in table (5).

Table (5) Multiple Linear Regression.

Data	P-value	std-Error	Estimate.	Min.	Max	Median.	R.squared
Cs.be.eng	0.6615	3.2076	-910.5	-103.4	99.364	2.769	0.008931
Cs.re.after	0.4547	1.3008	-2253.2				

The p-value based on Fisher's statistic is greater than 0.05 in table (5), indicating the presence of a variable and dependent of one since it has no effect on the data, as seen in figures (7).

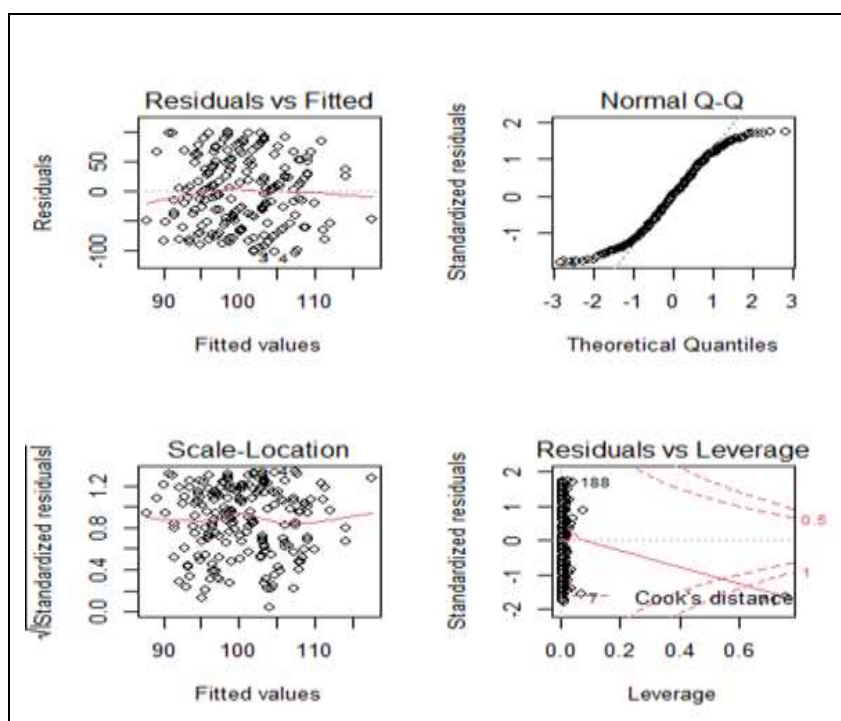
**Figure (7) Multiple Linear Regression for R.E.**

Figure (7) shows the results of tests of linear regression conditions. In the first diagram (Residuals vs Fitted), the validity of a linear relationship between sediments and estimators is investigated. Thus, it can be seen that the link does not adopt a precise shape, indicating that there is a flaw in the model that prevents this hypothesis from being realized.

The second diagram depicts the typical sediment distribution (Normal Q-Q). The sediments have committed to a normal distribution when data is taken around a straight line.

The third diagram (Scale Location) depicts the homogeneity of sediment variation, i.e., the data spreading process, in which the beginning and end of the



drawing must be comparable in shape, and when one end differs from the other, the heterogeneity of sediment variation is evident.

Outliers are shown in the fourth image (Residuals vs Leverage), because the smaller the sample size, the more imbalanced certain of the multiple linear regression requirements become.

Conclusions

In this study, statistical distributions and statistical software were used, and it was revealed to the following:

- 1-The Kolmogorov-Smirnov test revealed that the data in the sampling random analysis are distributed in a regular manner.
- 2-The R.E. of the spectrum data was found to follow a normal distribution in a Shapiro test, depending on the p. value.
- 3-In the Q.Q. Normal Test, the points in R.E are applied in a straight line, indicating that the data is homogeneous.
- 4-Using the linear regression test, it was determined that the analyzed random variables have no statistically significant correlation.
- 5-It was discovered that the analyzed samples (R.E) follow a normal distribution using the T-Test.
- 6-The third group in Pearson correlation was less than zero, while the rest of the groups were near zero, with the first and second groups having negative values, indicating that the investigated variables have a weak inverse link. The third group had a positive correlation value, indicating that the ^{137}Cs source's R.E. had increased.



Conflict of interests.

There are non-conflicts of interest.

References

- [1] G. E Knoll, "Radiation Detection and Measurement", Third Edition, John Wiley & Sons, Inc., (2000).
- [2] H.J. Arnika, "Essentials of Nuclear Chemistry", Fourth Edition, New Age International (p) Ltd., (2005).
- [3] J.E. Martin, "Physics for Radiation Protection, Third Edition, Published by Wiley-Viley-Vch GmbH & Co. KGaA., (2013).
- [4] J. Wiley, "Practical Gamma-ray Spectrometry", Second Edition, Gordon R. Gilmore Nuclear Training Services Ltd Warrington, UK, (2008).
- [5] J.L. Basdevant, J. Rich and M. Spiro, "Fundamentals in Nuclear Physics", Springer Science-Business Media, Inc., (2005).
- [6] G. R. Gilmore, "Practical Gamma-ray Spectrometry second Edition", Nuclear Training Services Ltd Warrington, UK, (2008).
- [7] P. Fanto, Y. Alhassid, and H. A. Weidenmuller, "Statistical-Model Description of γ Decay from Compound-Nucleus Resonances", Physical Review C, Vol. 101, No.1, p. 1-9, (2020).
- [8] M. G.-Fernandez, W. K. Wong, A. Tokuhito, K. Welter, A. M. Alhawsawi, H. Yang and K. Higley, "Isotope Identification Using Deep Learning: An Explanation", Nuclear Inst. and Methods in Physics Research, A, Vol. 988, No.16, p.608-615, (2021).
- [9] R. André, C. Bobin, J. Bobin, J. Xu and A. Vismes, "Metrological Approach of γ -Emitting Radionuclides Identification at Low Statistics: Application of Sparse Spectral Unmixing to Scintillation Detectors", Metrologia, Vol. 58, No. 1, p. 5011-5019, (2021).
- [10] M. R. Abonazel, "Advanced Statistical Techniques Using R: Outliers and Missing Data", Annual Conference on Statistics, Computer Sciences and Operations Research, College of Graduate Studies for Statistical Research, Cairo University, (2019).
- [11] M. R. Abonazel, "Statistical Analysis Using R", Annual Conference on Statistics, Computer Sciences and Operations Research, Egypt, (2014).
- [12] B. Yakir, "Introduction to Statistical Thinking, With R, Without Calculus", The Hebrew University, (2011).