Journal of Quality Measurement and Analysis *JQMA* 18(3) 2022, 61-70
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*

DETERMINE THE PARAMETERS FOR PHOTOELECTRIC EFFECT DATA USING CORRELATION AND SIMPLE LINEAR REGRESSION

(Menentusahkan Parameter untuk Data Kesan Fotoelektrik Menggunakan Kolerasi dan Linear Regressi Mudah)

SARATHA SATHASIVAM*, SALAUDEEN ABDULWAHEED ADEBAYO,
MURALY VELAVAN & JASON KNG WEI LIANG

ABSTRACT

Pearson's correlation coefficient, otherwise known as the product-moment correlation coefficient, a non-parametric process, is a very important concept in statistics, data science, and even in machine learning. It has gained tremendous acceptance in almost all fields and industries where data analysis is the business of the day. It helps to highlight the affinity between two variables whose behaviour might be entirely different, correlation coefficient is an indicator that shows whether such affinity is positive, negative, or none, when no linear relationship can be established between the variables. It is characterized by a numerical value that ranges between -1 and 1. These values serve as the indicators that determine the status of the relationship. In this research, we utilized the idea of correlation coefficient and simple linear regression on experimental data of photoelectric effects to determine the Planck constant, work function, and threshold frequency using MATLAB code.

Keywords: correlation coefficient; linear- regression; MATLAB code; Planck constant

ABSTRAK

Pekali korelasi Pearson, atau dikenali sebagai pekali korelasi momen produk, proses bukan parametrik, ialah konsep yang sangat penting dalam statistik, sains data dan juga dalam pembelajaran mesin. Ia telah mendapat penerimaan yang luar biasa dalam hampir semua bidang dan industri yang mana analisis data adalah perniagaan pada hari ini. Ia membantu untuk menyerlahkan pertalian antara dua pembolehubah yang tingkah lakunya mungkin berbeza sama sekali, pekali korelasi ialah penunjuk yang menunjukkan sama ada pertalian tersebut adalah positif, negatif atau tiada, apabila tiada perhubungan linear boleh diwujudkan antara pembolehubah. Ia dicirikan oleh nilai berangka yang berjulat antara -1 dan 1. Nilai ini berfungsi sebagai penunjuk yang menentukan status perhubungan. Dalam kertas kerja ini, kami menggunakan idea pekali korelasi dan regresi linear mudah pada data eksperimen kesan fotoelektrik untuk menentukan pemalar Planck, fungsi kerja, dan kekerapan ambang menggunakan kod MATLAB.

Kata kunci: pekali korelasi; regresi linear; kod MATLAB; pemalar Planck

1. Introduction

The discovery made by Heinrich Hertz that when the light of adequate frequency is shone on a metal surface and resulted in the emission of electrons from the surface of the metal was earlier termed Hertz effect, which was later changed to the photoelectric effect (Chakarvarti & Sharma 1988). Photoelectric effect is a vital phenomenon that manifests when electrically charged particles (electrons) are emitted within or from the surface of a material (usually metals) when electromagnetic radiation of an adequate (suitable) frequency is incident on metal (Marghany 2019). In other words, the emission of electrons or other free carriers when light strikes a metal is technically referring to photoelectric effect. The emitted electrons are known as photoelectrons. Photoelectric effect has usability in many areas of research and studies, majorly in electronic physics and modern physics. Recent studies have extended its applicability to chemistry, particularly electrochemistry and quantum chemistry (Spagnolo *et al.* 2020). This effect can be ascribed to the transfer of energy from incidental light to electrons.

Classical electromagnetic theory opines that a change in the intensity (brightness) of the light would result in a change of energy of the electrons released from the substance. In addition, the theory proposed that a sufficiently faint light will exhibit a delay between its initial illumination and the subsequent emission of an electron (Marghany 2019). Thus, neither of the two hypotheses is responsible for triggering electrons to move away from the surface of metals. It was, however, reported that emission of electrons will take place, at any frequency higher than the threshold frequency. The kinetic energy of the emitted electrons is inherent to the intensity of light but depends on the frequency of the incidental light (Sakho 2019). Thus, it justified that no matter what the intensity of the light is or the length of time of exposure to the material, as long as it is below the threshold frequency, no electron will be ejected from the material (Kloprogge & Wood 2020).

This phenomenon came with a complex scenario because of the perplexing twist about the nature of light-particle versus wavelike features of light, which were finally demystified by Albert Einstein in 1905 (Kloprogge & Wood 2020). The photoelectric effect is critically important in the development of modern physics and other research fields in astrophysics, technology, and materials sciences while serving as inspiration for many practical gadgets. Chodos *et al.* (2005) reported that photoelectric effect has a very interesting historical background, they detained it further that its history commenced when Heinrich Rudolf Hertz, a German Physicist, discovered in 1887 that electrodes exposed to UV light produced electric sparks more quickly. In the same light, another German Physicist Max Planck reported in 1900, that energy carried by electromagnetic waves could only be released as "packets" of energy whileconducting researching on black-body radiation.

Albert Einstein expedite the theory by his declaration that light energy is transported in discrete quantized packets (photons), He postulated that a photon penetrates the material and then transfers its energy to electrons, forcing electrons to acquire more energy and move at a high speed which eventually make them to escape from the surface of the material, the kinetic energy of an electron escaping from the surface of metal would reduce by the amount of work function (Kim *et al*. 2020). Planck's constant, often designated by h, is one of the fundamental physical constants in modern physics, it describes the quantum nature of energy as well as the behavioural state of particles at the atomic level while explaining how energy is quantized with respect to the spectral distribution of electromagnetic radiation (Checchetti & Fantini 2015). The numerical computation of the Planck constant (h) was initiated by Planck himself in 1900 where he stated the numerical value of h as 6.55×10^{-34} Js. A remarkable accomplishment was made in 1914 when an American physicist, Millikan, computed the value of the Planck constant experimentally using the photoelectric effect and presented the result of his experiment for h as $6.626\times10^{-34}Js$. Since then, numerous research has been conducted utilizing a variety of methods and techniques to determine the value of this quantity, the majority of which are based on the application of the photoelectric effect. In this research, we employed the idea of correlation coefficient and simple linear regression on experimental data of photoelectric effects to determine the value of Planck constant, work function, and threshold frequency using MATLAB code.

2. Mathematical Formulation

Einstein utilized Plank's quantum theory and derived a relation/equation for the photoelectric effect which is presently known as the Einstein Photoelectric equation whose success revolves around these assumptions (Abdelhady 2011).

- Light is made up of quanta or photons of energy, while each photon has an energy which is hv, where h is Planck's constant, and v is the speed of light.
- During the collision of a photon and an electron, the photon donates all its energy to the electron.
- A portion of the energy is utilized by the electron to leave the metal's surface, and the remaining portion is the kinetic energy after emission.
- The photoelectric work function (ϕ) of the metal is the minimal amount of energy needed for an electron to leave the metal surface.
- The residual energy $(hv-\phi)$ is the maximum kinetic energy for which a phoelectron is emitted.

Einstein photoelectric equation or simply photoelectric equation is written as

$$E_k = hv - \phi. \tag{1}$$

where E_k is the maximum kinetic energy of the photoelectron, h is Planck's constant, v is the frequency of incident light it is sometimes denoted, and ϕ is work function. The energy required to extricate an electron from the surface of a metal. The difference between the photon's energy and the surface's work function yields the net energy (Mcnichols *et al*. 2020). Expression of the work function is $\phi = hv_o$, where v_o is the threshold frequency. The threshold frequency v_o of the electromagnetic ray is defined as the minimum frequency which will lead to photoemission when light strikes the surface of a material, the relation in (1) can be equivalently rewritten as

$$E_k = hv - hv_o \tag{2}$$

3. Statistical Formulation

Many research works featured more than one variable, as such a researcher may be interested in getting a clear picture of the relationship between the variables under consideration. For instance, to track rainfall volume and plant growth rate, the number of eggs and numbers of layers in a poultry farm, soil erosion and rainfall volume over a given period of time. However, instead of analysing each variable independently as univariate data, it is more economical to garner pairs of data and come up with a robust way to characterize them as bivariate data, this is attainable by examining how one variable responds to changes in the values of the other variable, such response will assist the researcher to determine whether there is any relationship between the two variables under discussion or not. Such relation can graphically and mathematically depict viable information as well as affinity between the two variables. Therefore, the statistical connection between two variables is what is termed as correlation. It can be shown graphically as a scatterplot or as a chart. So that the pattern established can then be analysed.

Correlation coefficients constitute one of the numerous apparatuses used in measuring the degree of linear relationship between two variables, many correlations coefficient has been proposed in the past, such as simple correlation coefficient, population correlation coefficient, covariance correlation coefficient, Pearson's correlation coefficient, spearman's correlation coefficient, polychromic correlation coefficient are among frequently used correlation coefficient in statistical analysis of data (Akoglu 2018). Although, this research focused on the most frequently used correlation coefficient which is Pearson's correlation. The nature of the relationship that can be measured by the Pearson's correlation coefficient is monotonic since it, measures only the degree of responsiveness of one variable to the other. Thus, when one variable's value rises, so does the value of the other variable, or as one variable's value rises, the value of the other variable falls. In other word, higher (positive correlation) or lower (negative correlation) values of one variable frequently follow higher (one variable) or lower (the other variable) values of the other variable. However, Pearson product-moment is commonly designated as *'r'* and can be computed using (3) (Akoglu 2018).

$$r = \frac{\left[n\sum xy - \left(\sum x\right)\left(\sum y\right)\right]}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}}. \qquad (3)$$

The correlation coefficient *(r)* is sometimes regarded as the Pearson product-moment correlation coefficient (Gideon 2007). It is unitless and it is quantified by a numerical value between -1 and +1, a numerical value of zero (0) indicates there is no correlation between the variables while a numerical value of 1 or -1 often refers to as complete or perfect correlation and indicates strong correlation between the variables under consideration.

According to (Morrison 2014), simple linear regression allows us to predict response or dependent variable for a set of predictor values within the range of data through a phenomenon called interpolation. Model created using this method yields a mathematical equation of the form of $\hat{y} = \hat{m}x + \hat{c}$, where $\hat{m}$ is the gradient or slope and $\hat{c}$ is the *y*-intercept of the regression line. The gradient and the interception can be written in statistical form as shown in (4) and (5) respectively.

$$\hat{m} = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}. \qquad (4)$$

$$\hat{c} = \frac{\left(\sum y\right)\left(\sum x^2\right) - \left(\sum x\right)\left(\sum xy\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}. \qquad (5)$$

However, it is very important to estimate the two standard errors of estimation of the regression line's slope and intercept. The standard error of a regression slope $\hat{m}$ denoted by $S_{\hat{m}}$ and that of interception $\hat{c}$ is designated by $S_{\hat{c}}$, and can be computed using (6) and (7) respectively.

$$S_{\hat{m}} = \sqrt{\frac{\left(\frac{1}{n-1}\right)\sum (y - \bar{y})^2}{\sum (x - \bar{x})^2}}. \qquad (6)$$

where

n = total sample size,
y = actual value of the response variable,
$\overline{y}$ = predicted value of the response variable,
x = actual value of the predictor variable,
$\overline{x}$ = predicted value of the predictor variable.

Similarly, the standard error of a regression line intercept can be calculated using the relation

$$S_{\hat{c}} = \sqrt{\left[\left(\frac{1}{n-2}\right)\sum (y-\overline{y})^2\right]\left[\left(\frac{1}{n}+\frac{\overline{x}^2}{\sum (x-\overline{x})^2}\right)\right]} \tag{7}$$

Eq. (6) and Eq. (7) can be equivalently written as (8) and (9).

$$S_{\hat{m}} = \sqrt{\left(\frac{1}{n-2}\right)\frac{\sum (y-\overline{y})^2}{n\sum x^2 - \left(\sum x\right)^2}} \tag{8}$$

$$S_{\hat{c}} = \sqrt{\left(\frac{1}{n-2}\right)\frac{\sum (y-\overline{y})^2 \sum x^2}{n\sum x^2 - \left(\sum x\right)^2}} \tag{9}$$

The smaller the value of (8) and (9) the lower the variability around the coefficient estimate of the regression and the better the solution obtained.

4. Methodology

In this research, we utilized a methodology which revolves around associating relevant variables to each other, by comparing the quantity $E_k = hv - hv_o$ to $\hat{y} = \hat{m}x + c$ we have $\hat{y} = E_k$, $x = v$, $\hat{m} = h$, $\hat{c} = -\phi = -hv_o$ Since we are dealing with photoelectric effect and our main target is to compute the value of Planck constant, work function and the threshold frequency. we utilized correlation coefficient and simple linear regression. Representing stopping potential in the *x*-axis and frequency in *y*-axis. Relevant information was generated using correlation coefficient and the simple linear regression. The dataset obtained from Step by Step Science (2019) is shown in the Table 1, while experimental dataset obtained from André and Brito (2014) is shown in Table 2.

The following relations were used for our comparison, Planck constant (h) can be set as $h = \hat{m}$, error in calculation Planck constant $= S_{\hat{m}}$, work function $\phi = -\hat{c}$, Error in calculating work function $= S_{\hat{c}}$, and finally, the threshold frequency was computed using (10).

$$v_o = -\frac{\hat{c}}{\hat{m}} = \frac{\left(\sum x\right)\left(\sum xy\right) - \left(\sum y\right)\left(\sum x^2\right)}{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum xy\right)}. \tag{10}$$

Table 1: Data of Stopping potential and wavelength of light (Step by Step Science, 2019)

Wavelength, $\lambda(nm)$	403	446	569	611	738
Stopping potential, $U(V)$	1.80	1.50	0.90	0.75	0.40

Table 2: Data of Stopping potential and wavelength of light (André & Brito, 2014)

Wavelength, $\lambda(nm)$	623	586	567	467
Stopping potential, $U(V)$	1.78	1.90	2.00	2.45

5. Flow Chart

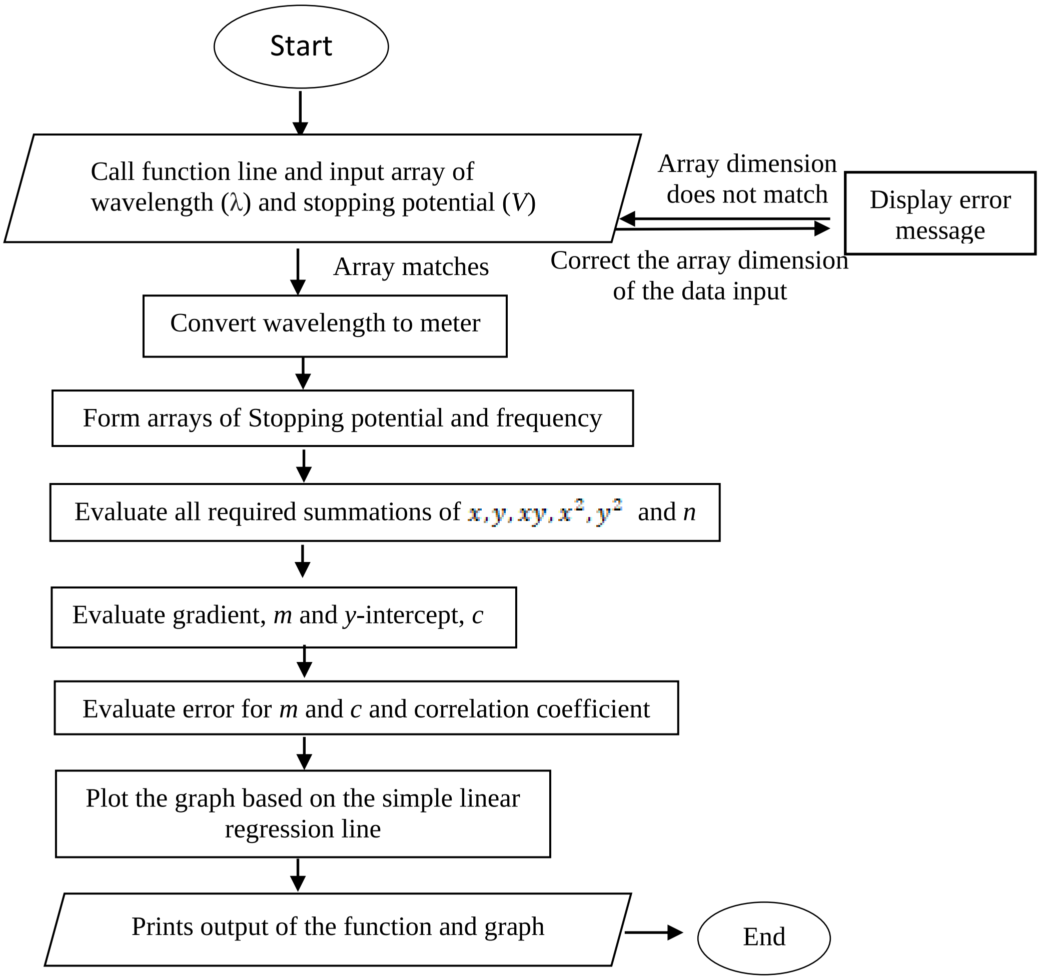

Figure 1: The flowchart showing the build-up processes

Step 1: An array of values representing the wavelengths (*nm*) and array of stopping potential (*V*) must be specified. The dimension of the two parameters must be the same so that, for each wavelength there is a corresponding stopping potential.

Step 2: To have a consistent representation, convert all values to *SI* units. Therefore, the nanometer, which is the unit of the wavelength, must be appropriately converted to meters.

Step 3: Compute an array of frequencies using the relationship that connect, the speed of light, wavelength, and frequency. Similarly, the relation between the maximum kinetic energy and stopping potential is used to generate the array of stopping criterial in joule *(J)*. therefore we represents stopping potential and frequencies in the *y* - axis and *x* -axis respectively.

Step 4: Identify the values for *x* and *y* that corresponds to each frequency and corresponding stopping potential respectively. Computation of terms needed in our relation such as x, y, xy, x^2, y^2 .

Step 5: By using (4) and (5), compute the value *m* and *c*, which represents the slope and the *y*-intercept of the graph respectively.

Step 6: Using (8) and (9) compute the standard error of estimation of $\hat{m}$ and $\hat{c}$ which are designate as $S_{\hat{m}}$ and $S_{\hat{c}}$ respectively.

Step 7: To vitalize the graphical representation of the datasets under consideration, plot the values of *x* against value of *y* on the cartesian plane.

Step 8. Analysis of the features of the graph obtained in step 7, in term of the correlation coefficient, computation of wavelength, frequencies, Planck constant and so on.

6. Results and Discussion

Computation from our MATLAB code using the relation (4, 5, 6, 7 & 9), gave results in Table 3 for dataset in Table 1 and results in Table 4 for experimental data in Table 2. All the calculations were performed using *S I* units.

Table 3**:** Results from the datasets in Table 1

$m = 6.6234\times10^{-34}\,Js.$
$c = -2.0525\times10^{-19}\,m.$
$r = 1.0000.$
Planck constant $(h) = 6.6234\times10^{-34}\,Js.$
Error in computing Planck constant $= 7.0647\times10^{-37}\,Js.$
Work Function$(W_o) = 2.0525\times10^{-19}\,J.$
Error in computing work function $= 4.1084\times10^{-22}\,J.$
Threshold frequency $= 3.098\times10^{14}\,Hz.$

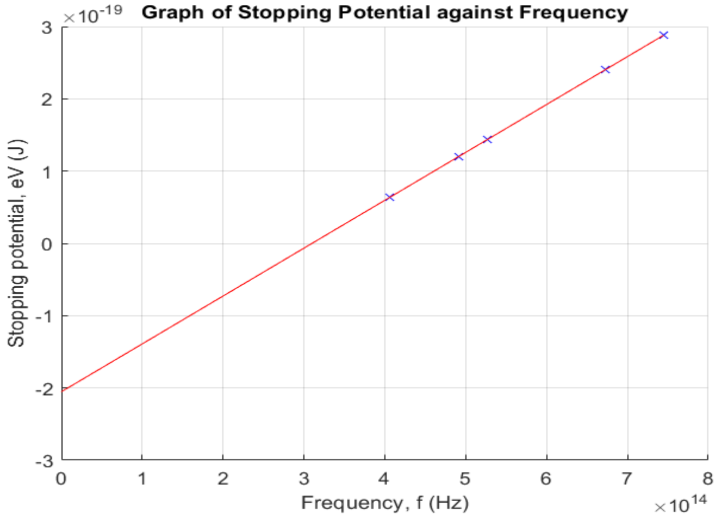

Figure 2: The graph of the stopping voltage against the frequency for the dataset

Table 4**:** Results obtained from the datasets in Table 2

$m = 6.6566 \times 10^{-34}\, Js.$
$c = -2.0525 \times 10^{-20}\, J.$
$r = 0.9991.$
Planck constant $(h) = 6.6566 \times 10^{-34}\, Js.$
Error in computing Planck constant $= 2.006 \times 10^{-35}\, Js.$
Work Function$(W_o) = 3.5084 \times 10^{-20}\, J.$
Error in computing work function $= 1.0926 \times 10^{-20}\, J.$
Threshold frequency $= 5.2706 \times 10^{13}\, Hz.$

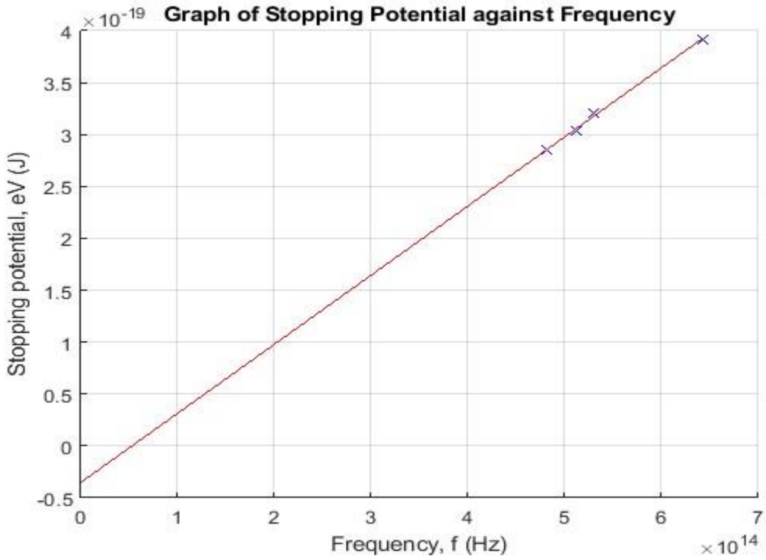

Figure 3: The graph of the stopping voltage against the frequency for the dataset

From the relation of Stopping potential and frequency, we expect the slope of the graph to give us the value of the Planck constant, which can then be used to compute the work function and threshold frequency. Figures 2 and 3 show the graphs obtained by graphing the value of stopping potential *(V)* against the frequency *(f)* from our two datasets. Thus, a key factor in establishing a quality work function and threshold frequency depends largely on the goodness of the slope of the graphs (Planck constant), which can be assessed by comparing it to the universal Planck constant.

As shown in Table 3, Planck constant was estimated to be $6.6234\times10^{-34}\,\mathrm{Js}$. This value is actually very close to universal Planck constant's well-known value of $6.626\times10^{-34}\,\mathrm{Js}$. It can be observed from figure 2 that, the regression line fit the data properly. Similarly, computed value of the Planck constant is the gradient of our regression line. Additionally, the standard error of estimation of Planck constant is $2.006\times10^{-35}\,\mathrm{Js}$. This error is also very small compare to computed Planck constant value. The work function, which is obtained as $2.0525\times10^{-19}\,\mathrm{J}$, is the negative value of the *y*-intercept of the regression line. The standard error of estimation of work function, which was obtained as 4.1084×10^{-22}.Therefore, we asserted high degree of assurance that the work function value obtained is accurate since the estimation error is very small as well. However we also arrived at the conclusion the light ray's threshold frequency of 3.0988×10^{14} Hz is extremely accurate. The correlation coefficient is computed to be 1, this reaffirmed the goodness of the regression line in fitting the data well. Although, this is an uncommon result, since it shows a perfect correlation between the two variables. These kinds of correlation are called perfect correlations, and are uncommon, instead, value of r that are roughly close to absolute value of one can as well be considered as suitable indication of correlation between two variables.

In Table 4, the Planck constant was obtained as $6.6566\times10^{-34}\,\mathrm{Js}$. This figure is not as close as the one obtained in the Table 3. This can be attributed to some data points that did not fit well. in other word, the regression line did not fit the experimental data properly, an obvious reason for *r=0.9991,* although it is very close to 1. but it is an indication that some data points did not fits well this can be observed easily from figure 3, the standard error of estimation of Planck constant was found to be 7.0647×10^{-37}. This error is also considerably very small in relation to the estimated Planck constant value. Thus, Planck constant value of such nature will hinder the accuracy of the work function and computed threshold frequency.

7. Conclusion

Several academic papers on the photoelectric effect have explained the relationship between the Planck constant, the work function, and the threshold frequency. In the same light, Important information regarding correlation coefficient, statistical and numerical aspects of simple linear regression were equally explored. To make this research self-contained, pertinent equations and relations were highlighted and simplified. The goal of the study is to use MATLAB function code to calculate the Planck constant, work function, and threshold frequency from photoelectric effect data. We built the MATLAB function codes to address both stimulated data and experimental data.

Our approach has proven to be very effective because it captured the trend in data points by showing a better result for the data that was free from error by producing a very good Planck constant value, and a relatively fair result for the data that had few outliers and did not fit well. Similarly, it shows high capability in producing quick and accurate results for the photoelectric effect data, producing quick results and values for a particular material or metal. This will help in determination of material suitability in product selection. However, since it is a function code, it can be used multiple times to generate and compare regression lines for multiple materials at the same time, allowing effective and quick decision-making in the choice of material for application.

References

Abdelhady S. 2011. Comments on Einstein's explanation of electrons, photons, and the photoelectric effect. *Applied Physics Research* **3**(2): 230-240.

Akoglu H. 2018. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine* **18**(3): 91-93.

André F. & de Brito André P.S. 2014. Classroom fundamentals: Measuring the Planck constant. on Science in School. http://www. scienceinschool. org/2014/issue28/planck (12 June 2022).

Chakarvarti S.K & Sharma. B.L. 1988. Determination of Planck's constant using the photoelectric effect. *Physics Education* **23**(4): 249.

Checchetti A. & Fantini A. 2015. Experimental determination of Planck's constant using light emitting diodes (LEDs) and photoelectric effect. *World Journal of Chemical Education* **3**(4): 87-92.

Chodos A.J. Ouellette, & E. Tretkoff. 2005. Einstein predicts stimulated emission. *American Physical Society* **14**(8).

Gideon R.A. 2007. The correlation coefficients. *Journal of Modern Applied Statistical Methods* **6**(2): 517-529.

Kim E., Kim H., Lee J. & Lee G. 2020. How Modern Physics Textbooks Explain Intensity of Light in Photoelectric Effect. *Journal of Science Education* **44**(1): 112-121.

Kloprogge J.T. & Wood B.J. 2020. *Handbook of Mineral Spectroscopy. Volume 1: X-ray Photoelectron Spectra*. Amsterdam: Elsevier.

Makowski A.J. 2000. A century of the Planck constant. *Physics Education* **35**(1): 49.

Marghany M. 2019. *Synthetic Aperture Radar Imaging Mechanism for Oil Spills*. Cambridge, MA: Gulf Professional Publishing.

Mcnichols C., Gaudreau I., Wang M., Yuan M., Meng Q. & Machado A. 2020. Measuring the work function of metals through the Photoelectric effect. Department of Physics, Engineering Physics and Astronomy: PHYS 350 General Physics Laboratory Collection. Queen's University.

Morrison F.A. 2014. Obtaining uncertainty measures on slope and intercept of a least squares fit with Excel's LINEST. Michigan Technological University.

Sakho I. 2019. *Introduction to Quantum Mechanics 1: Thermal Radiation and Experimental Facts Regarding the Quantization of Matter*. NJ: John Wiley & Sons, Inc.

Smith G. 2015. *Essential Statistics, Regression, and Econometrics*. 2nd Ed. San Diego, CA: Academic press.

Spagnolo G.S., Postiglione A. & De Angelis I. 2020. Simple equipment for teaching internal photoelectric effect. *Physics Education* **55**(5): 055011.

Step by Step Science. 2019. Photoelectric effect (7 of 8) determining Planck's constant. https://www.youtube.com/watch?v=3AlqTr1sgnM (10 May 2022).

School of Mathematical Sciences,
Universiti Sains Malaysia, Penang,
11800, USM,
Malaysia
E-mail: saratha@usm.my, salaudeenadebayo@gmail.com, jasonkng99@student.usm.my*

School of General and Foundation Studies,
AIMST University,
08100 Bedong, Kedah,
Malaysia
E- mail: muraly@aimst.edu.my

Received: 4 October 2022
Accepted: 4 November 2022

*Corresponding author