Turkish Journal of Agriculture - Food Science and Technology, 12(s2): 2234-2241, 2024
DOI: https://doi.org/10.24925/turjaf.v12is2.2234-2241.6931

Turkish Journal of Agriculture - Food Science and Technology

Available online, ISSN: 2148-127X | www.agrifoodscience.com | Turkish Science and Technology Publishing (TURSTEP)

On The Pros and Cons of Using Excel for Regression Analysis

Sencer Buzrul[1,a,*]

[1]*Necmettin Erbakan University, Department of Food Engineering, Meram/Konya, Türkiye*
[*]*Corresponding author*

ARTICLE INFO

Research Article

Received : 10.06.2024
Accepted : 16.07.2024

Keywords:
Curve fit
Goodness-of-fit
Spreadsheet
Software
Summary output

ABSTRACT

Excel is a useful and powerful software for simple regression analysis without any programming skills and that is why, it is widely preferred by the undergraduate students from various areas such as chemistry, biology and agriculture as well as many engineering disciplines such as chemical engineering, food engineering and bioengineering. Parameter values and coefficient of determination (R^2) can be easily obtained together with the graphical representation for those models exist in Excel such as linear and exponential models. It is also possible to visually examine the model fit and experimental data together on the same graph. For linear models (linear in parameters) Excel Add-In Data Analysis-Regression tool creates a summary output, and parameter estimates, parameter uncertainties, adjusted R^2 (R^2_{adj}) and root mean square error (RMSE) values can be found even for the models that do not exist in Excel. For nonlinear models (nonlinear in parameters) Excel Add-In Solver tool can be used to obtain parameter estimates (but not uncertainties), and R^2, R^2_{adj} and RMSE can be calculated manually. Despite these advantages, there are some shortcomings of Excel for regression analysis. For linear models with no-intercept Excel reports the incorrect and overoptimistic R^2 and also reports incorrect and overpessimistic R^2_{adj}. Excel has also some nonlinear models such as exponential and power models in it; however, Excel computes linear parameter estimates for those nonlinear models and again optimistic R^2 is calculated. This paper aims (i) to show these inaccuracies with their reasons by using published data, (ii) how to obtain the correct results, (iii) to warn the instructors who would use Excel for regression analysis in class. Excel users (instructors, students, professionals in any field) should be aware of the pitfalls when using Excel for regression.

[a] *sencer.buzrul@erbakan.edu.tr* *https://orcid.org/0000-0003-2272-3827*

Introduction

Software programs are being used not only for data analysis but also for teaching purposes. Microsoft Excel is very popular for data collection and storage and it is the first choice for especially undergraduate students as well as instructors. Moreover, professionals working in any field and dealing with data also use Excel. This is mostly due to Excel's versatility, availability and ease of use (Serment-Moreno, 2021). It seems that Excel is one of the best software for teaching statistical analysis, graphic data presentation and data management (Rubin & Edams, 2015). Regression is a widely used technique in almost all areas such as chemistry, biology, agricultural science, biotechnology, chemical engineering, food engineering and bioengineering. Excel also provides regression analysis without any knowledge of a programming language which is very advantageous for undergraduate students as well as graduates from the aforementioned disciplines. Despite all these positive features, earlier versions of Excel were criticized by statisticians. Keeling & Pavur (2011) compared six spreadsheet packages including Excel 2007 and 2010 and noted that Excel 2010 had significant improvement especially on the numerical accuracy of the statistical distribution tests compared to Excel 2007. Besides, Excel 2010 and Gnumeric (a spreadsheet package) performed the best out of all packages on regression datasets. Although Microsoft partially improved some aspects of Excel throughout the years (Mélard, 2014), there are still some problems in Excel especially for regression analysis.

Excel users should be aware of some false results reported by Excel in regression analysis and should know how to find the correct results. Therefore, the aim of this paper was to (i) discuss some of the positive and negative features of Excel for regression analysis, (ii) show incorrect results calculated by Excel with some examples and (iii) also show how to correct those results again by using Excel.

Some Advantages of Excel for Regression Analysis

It is possible to sketch the *x*-*y* data in Excel easily and to fit some equations (such as linear and exponential equations) to data which are available in Excel. Users can fit the model and obtain the parameter estimates as well as coefficient of determination (R^2) on the graph. However, uncertainties in parameters (standard errors or confidence intervals) and other goodness-of-fit indices such as adjusted R^2 (R^2_{adj}) and standard error of the model also known as root mean square error (RMSE) cannot be obtained by this way. This is a major disadvantage because parameter uncertainties are as important as the parameters themselves (Denton, 2000; van Boekel, 1996). Excel has a great solution for this: Data Analysis ToolPak. This Excel Add-In has several applications and regression is one of them. Users who wish to use regression application follow the path: Data > Data Analysis > Regression in Excel and enter the *x* and *y* data to obtain a "Summary Output".

The summary output provides parameter estimates, standard error, 95% upper and lower limits (by default) or 99% upper and lower limits (users should write it manually, or should write any other desired percentage), correlation coefficient between the experimental data and the fitted model (named as multiple R in the summary output), R^2, R^2_{adj} and RMSE values. Any linear model (linear in parameters) can be analyzed this way whether the model exists in Excel or not.

Consider a simple linear regression in the form of:

$$y = a \cdot x + b \tag{1}$$

where *y* is the dependent variable, *x* is the independent variable, *a* (slope) and *b* (intercept) are the model parameters. This model (exist in Excel) can directly be fitted after the graph is sketched and this provides users to visualize or observe the straight line and data together on the graph. Then users can utilize Data Analysis > Regression tool and can enter the *x* (shown as Input X Range in Regression tool) and *y* (shown as Input Y Range in Regression tool) and have the results.

Now let us consider a quadratic polynomial:

$$y = A \cdot x^2 + B \cdot x + C \tag{2}$$

Since this model also exists in Excel (users have options for higher order polynomials such as third- or fourth-order), after sketching the data users can select this model to observe the model and data together on the graph. Parameter values and R^2 are also displayed. Moreover, the model is linear in parameters and therefore linear regression (polynomial or curvilinear regression) can be used to calculate the parameter values and goodness-of-fit statistics in Excel. However, it is expected that users should calculate x^2 in the cells next to *x* before to use Data Analysis > Regression tool and enter not only the *x* values but also x^2 values for the dependent variable since there is only one space for *x* (shown as Input X Range in Regression tool). This is a major drawback because users need to make extra calculations which is not a limitation for other statistical software packages because they are designed solely for this purpose. Nevertheless, by doing this, Excel will provide the summary output and all necessary information.

As a last example, let us consider Van Deemter equation:

$$y = A \cdot x + B/x + C \tag{3}$$

This model is also linear in parameters (and hence linear regression can be used) but does not exist in Excel. The model is used to define chromatography data where the dependent variable is plate height in mm and the independent variable is flow rate in mL/min (Harris, 1998). After the calculation of 1/x next to x, both x and 1/x can be entered in X, and y can be entered in Y section in regression tool and this provides the summary output. Since the model is not in Excel, to observe the model fit and data together on the same graph, users should use the parameter values given in the summary output and calculate the model estimates for each *x* and then add the model fit to the graph. All of those (calculations and model addition) can be done in Excel (Leylak et al., 2020); however, depending on Excel experience of the users this would take one to several minutes (3-5 min). This is another limitation of using Data Analysis ToolPak for regression analysis over other statistical software packages.

For nonlinear models (nonlinear in parameters) Data Analysis > Regression tool is useless since it is designed for linear models but Solver Add-In (Data > Solver) can be used (Brown, 2001; Kemmer & Keller, 2010; Yurdakul et al., 2020). This is an iterative procedure and initial parameter estimates should be entered by the user. Unfortunately, only the parameter estimates can be obtained, not the uncertainties by this procedure. Moreover, all goodness-of-fit statistics should be calculated by the users manually. SolverAid (De Levie, 2012) which is an Excel macro can be used to obtain the standard errors of the parameters but it seems that not too many people use this macro (van Boekel, 2022). On the other hand, statistical software packages used for nonlinear regression not only estimate the parameters but also standard error of the parameters. Some packages can even calculate the upper and lower confidence limits.

Some Examples of, and Some Problems with, the Use of Excel for Regression Analysis

Linear Models with No-Intercept

Excel can be safely used for simple linear regression [Eq.(1)] as said above. Consider the data given in Table 1. Independent variable (*x*) is concentration in μg/mL and dependent variable (*y*) is corrected absorbance (595 nm) in AU.

The data can be sketched and a straight line can be easily fitted in Excel (Figure 1).

Table 1. Absorbance data taken from Harmer & Hill (2021)

x	y
0	0.000
40	0.189
80	0.253
120	0.385
160	0.603
200	0.680

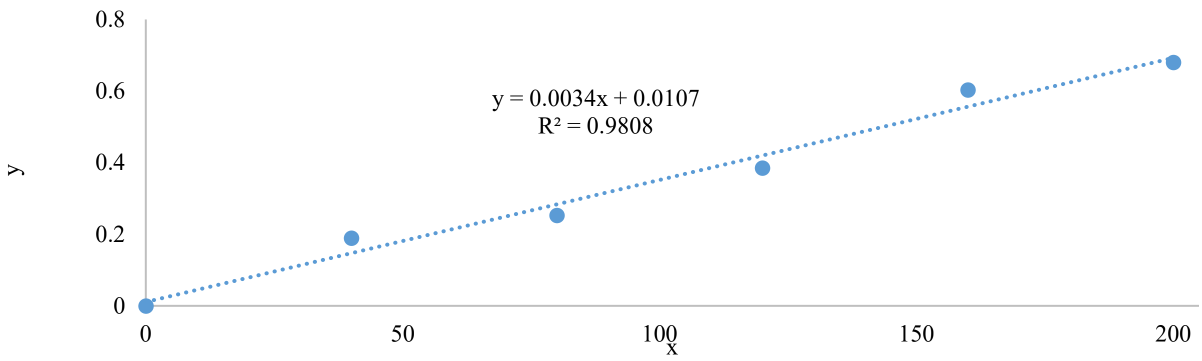

Figure 1. Fit of the linear model [Eq.(1)].

	A	B	C	D	E	F	G	H	I	J
1	SUMMARY OUTPUT									
2										
3	*Regression Statistics*									
4	Multiple R	0.9904								
5	R Square	0.9808								
6	Adjusted R Square	0.9760								
7	Standard Error	0.0399								
8	Observations	6								
9										
10	ANOVA									
11		*df*	*SS*	*MS*	*F*	*Significance F*				
12	Regression	1	0.3256	0.3256	204.75	0.000138572				
13	Residual	4	0.00636	0.00159						
14	Total	5	0.33195							
15										
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99.0%*	*Upper 99.0%*	
17	Intercept (b)	0.01067	0.02886	0.37	0.730419927	-0.069	0.091	-0.122	0.144	
18	Slope (a)	0.00341	0.000238	14.31	0.000138572	0.00275	0.00407	0.00231	0.00451	
19										

Figure 2. Summary output of the linear model [Eq.(1)] given by Excel.

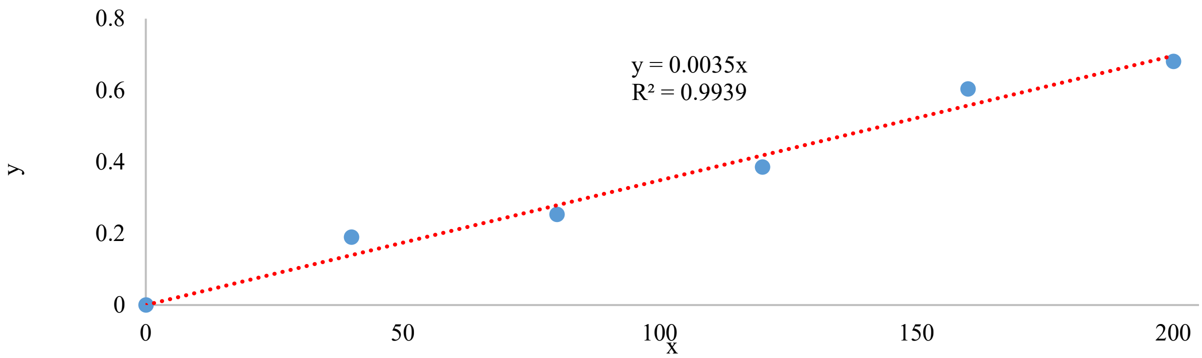

Figure 3. Fit of the linear model with no-intercept [Eq.(4)].

Parameter values as well as R^2 were found (Figure 1); however, parameter uncertainties and other goodness-of-fit indices cannot be obtained. To have those, Data Analysis > Regression in Excel was used x and y data were entered to obtain a “Summary Output”. This output is shown in Figure 2.

It was now possible to observe not only the parameter estimates but also their standard error values and their 95% and 99% lower and upper confidence limits. Moreover, R^2 (also found in Figure 1), R^2_{adj} and RMSE values were also obtained. A careful inspection of Figure 2 revealed that parameter b (intercept) was statistically insignificant ($p > 0.05$ or $p > 0.01$). Therefore, it might be better to repeat the regression without an intercept (data in Table 1 also supported this because $x = 0 \rightarrow y = 0$). In this case Eq.(1) becomes:

$$y = a \cdot x \tag{4}$$

Again, linear trendline was added but this time intercept was set to zero (Figure 3).

Slope (a) and R^2 were displayed on the graph. Note that the model without the intercept ($y = a{\cdot}x$) had higher R^2 than the model with intercept ($y = a{\cdot}x + b$) – see Figure 1 and Figure 3. To find the parameter uncertainty and other goodness-of-fit indices “Data Analysis > Regression” application was used once more (this time “Constant is Zero” option was selected.) and the summary output was obtained (Figure 4).

	A	B	C	D	E	F	G	H	I	J
1	SUMMARY OUTPUT									
2										
3	*Regression Statistics*									
4	Multiple R	0.9969								
5	R Square	0.9939								
6	Adjusted R Square	0.7939								
7	Standard Error	0.0363								
8	Observations	6								
9										
10	ANOVA									
11		*df*	*SS*	*MS*	*F*	*Significance F*				
12	Regression	1	1.0674	1.0674	811.3618	9.04E-06				
13	Residual	5	0.0066	0.0013						
14	Total	6	1.0740							
15										
16		*Coefficients*	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99.0%*	*Upper 99.0%*	
17	Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
18	Slope (a)	0.00348	0.00012	28.48	9.99E-07	0.00317	0.00380	0.00299	0.00398	
19										

Figure 4. Summary output of the linear model with no-intercept [Eq.(4)] given by Excel.

Table 2. Sum of squared error (SSE) and sum of squared total (SST) calculations for the linear model with no-intercept [Eq.(4)] and for the data given in Table 1

x	y	y_{model} (0.00348·x)	$(y - y_{model})^2$	$(y - y_{mean})^2$
0	0.000	0.0000	0.000000	0.1237
40	0.189	0.1392	0.002480	0.0265
80	0.253	0.2784	0.000645	0.0097
120	0.385	0.4176	0.001063	0.0011
160	0.603	0.5568	0.002143	0.0632
200	0.680	0.6960	0.000256	0.1078
y_{mean} = 0.35167			SSE: 0.006587	SST: 0.332

In the summary output given in Figure 4, the same R^2 value shown on the graph (Figure 3) was found. Users generally satisfy by these results because as said earlier higher R^2 was found [R^2 = 0.9808 for Eq.(1) and R^2 = 0.9909 for Eq.(4)] and the sole parameter in Eq.(4) was also statistically significant ($p \leq 0.05$ or $p \leq 0.01$); however, it is awkward that R^2_{adj} was much smaller than R^2 (Figure 4). Normally, R^2_{adj} is almost always smaller than R^2 but, it is expected that these two should be close to each other. Moreover, R^2 (both shown in Figure 3 and Figure 4) was rather optimistic and higher than the actual R^2 – see below.

R^2 can be calculated by using the following formula:

$$\mathrm{R}^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} \quad (5)$$

where SSE is sum of squared error i.e., sum of the difference between the (experimental) data and the model fit and SST is the sum of squared total i.e., sum of the difference between the (experimental) data and the mean or average of the data. These values were calculated by using Excel for Eq.(2). Results of the calculations are shown in Table 2.

It was now possible to calculate R^2 by using SSE and SST in Table 2 for Eq.(4) i.e. $y = 0.0348 \cdot x$

$$\mathrm{R}^2 = 1 - \frac{0.006587}{0.332} = 0.9802 \quad (6)$$

This R^2 value was smaller than the one reported by Excel in Figure 3 and Figure 4 which was 0.9939. A question arises here: How Excel could calculate different R^2 than the real R^2?

Normally SST should be calculated as:

$$\mathrm{SST} = \sum (y - y_{mean})^2 \quad (7)$$

However, Excel for no-intercept model calculates SST as:

$$\mathrm{SST} = \sum (y - 0)^2 \quad (8)$$

In other words, Excel sets the mean value as zero in regression through origin and this generally ends up with higher SST and hence higher R^2 – see Eq.(5). SST value calculated by using Eq.(8) was found as 1.074 which was exactly the same value in Figure 4. However, SST was found as 0.332 by using Eq.(7) (Table 2).

The formula of R^2_{adj} is given below:

$$\mathrm{R}^2_{\mathrm{adj}} = 1 - \frac{n-1}{n-p} \times (1 - \mathrm{R}^2) \quad (9)$$

where n is the number of data points and p is the number of parameters in the model. According to Eq.(9), R^2_{adj} and R^2 can only be equal to each other in two cases: (i) if $R^2 = 1$ then $R^2_{adj} = 1$ which means $R^2 = R^2_{adj}$. However, $R^2 = 1$ means perfect fit and this is not likely possible for any experimental data such as biological or chemical data. (ii) if there is only one parameter in the model then $R^2 = R^2_{adj}$. Other than these two cases R^2_{adj} should be smaller than R^2.

Since Eq.(4) had only one parameter R^2 should be equal to R^2_{adj}; however, it was totally a different value (Figure 4). Therefore, R^2_{adj} too should be calculated, but in our case, there was no need for another calculation because we already calculated and found R^2 value as 0.9802 which was also the value of R^2_{adj}.

RMSE can be calculated by using the following formula:

$$\mathrm{RMSE} = \sqrt{\frac{\mathrm{SSE}}{n-p}} \tag{10}$$

As said earlier there was only one parameter (p) in the model [Eq.(4)] and number of data points (n) was 6 (Table 1), and SSE was calculated as 0.006587 (Table 2). If these values were integrated into Eq.(10), RMSE can be found as 0.0363. Note that both SSE and RMSE values were same as the ones calculated by Excel (Figure 4). In conclusion, RMSE (standard error of the model) was the sole correct goodness-of-fit statistic in regression statistics table shown in Figure 4 with no-intercept model.

Now consider Eq.(2) with no-intercept, i.e.:

$$y = A \cdot x^2 + B \cdot x \tag{11}$$

If Data Analysis > Regression application of Excel is used for Eq.(11), R^2 and R^2_{adj} values would be again incorrect (RMSE would be correct) in the summary output and they should be recalculated by the user, but this time users should make two calculations: one for R^2 and one for R^2_{adj} because number of parameters in Eq.(11) is not one. Furthermore, the model should be added to the graph manually by the users to visualize the model fit and the experimental data on the same sketch.

Nonlinear Models

There are some nonlinear models such as exponential and power models in Excel; however, the trend lines of these models do not represent the best fit parameter estimates obtained from nonlinear regression. Excel linearizes these models by logarithmic transformations and computes parameter values obtained from linear regressions (Dolan & Mishra, 2013; Yurdakul et al., 2020). Moreover, R^2 value given by Excel for those models are also incorrect.

Let us give an example to show these shortcomings by using the exponential model in Excel which is in the form of:

$$y = G \cdot e^{-H \cdot x} \tag{12}$$

where y is the dependent variable, x is the independent variable, G and H are the model parameters.

As an example, consider the data given in Table 3. Original data were published by Halabi et al. (2020); however, data were directly taken from https://github.com/TinyvanBoekel/IDJ (van Boekel, 2022). Independent variable (x) is time in minutes and dependent variable (y) is concentration of α-lactalbumin in mg/L.

Table 3. Denaturation data of α-lactalbumin taken from van Boekel (2022)

x	y
0.0	1.288
1.1	1.214
3.1	1.088
5.1	0.969
7.1	0.844
9.1	0.769
14.1	0.596
19.1	0.389
24.1	0.279
29.1	0.295
39.1	0.135
49.1	0.122
59.1	0.066

Figure 5 shows the model fitting, model equation and R^2 computed by Excel. When Eq.(12) was fitted to the data G = 1.2002, H = 0.05 and R^2 = 0.9926. If the exponential model was linearized by logarithmic transformation:

$$\ln y = \ln G - H \cdot x \tag{13}$$

or

$$y' = G' - H \cdot x \tag{14}$$

where $y' = \ln y$ and $G' = \ln G$. It was possible to perform simple linear regression by using Eq.(13) or Eq.(14). Therefore, if the natural logarithm of y values given in Table 3 were calculated and then ln$y$ vs. x was plotted, Excel reported the results given in Figure 6.

Parameter H could be directly obtained from the linear regression as 0.0505 which has the same value as exponential model's parameter H (Note that it was found as 0.05, the difference was due to the way of reporting the digits in Excel) and ln$G$ = 0.1825 which means G = exp(0.1825) or G = 1.2002. In brief, both models i.e., Eq.(12) and Eq.(13) resulted in the same parameter estimate. This would lead to a wrong conclusion as there was no difference between linear and nonlinear regression results. As said before Excel linearizes the nonlinear equation and obtains the parameters via linear regression (not nonlinear regression!) and computes those parameters in the nonlinear model. Therefore, parameter values given in Figure 5 are not the best-fit parameters.

In order to obtain the parameters via nonlinear regression [Eq.(12)], Solver in Excel was used. The parameter values and R^2 were found as: G = 1.2893, H = 0.0572 and R^2 = 0.9957. The parameter values were close but not identical and so was R^2. Fits of both models (linear and nonlinear parameters estimates) are shown in Figure 7.

Visually results of non-linear regression were better because the orange lines passed closer to the data (blue circles) especially for the first four data points (Figure 7). Furthermore, R^2 of the nonlinear regression was higher than that of linear regression indicating a better fit, but unfortunately, there is another pitfall in exponential model: Excel reports the wrong and optimistic R^2 for its exponential model (Even if this was the case, the nonlinear regression produced better R^2 value).

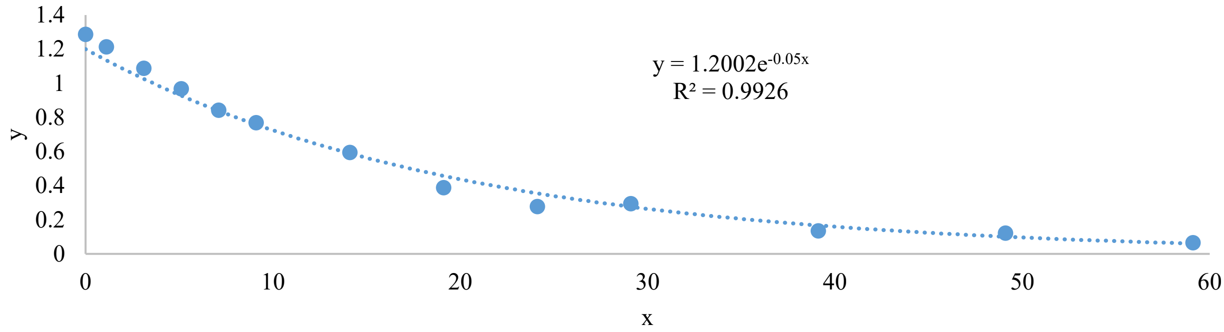

Figure 5. Fit of the exponential model [Eq.(12)].

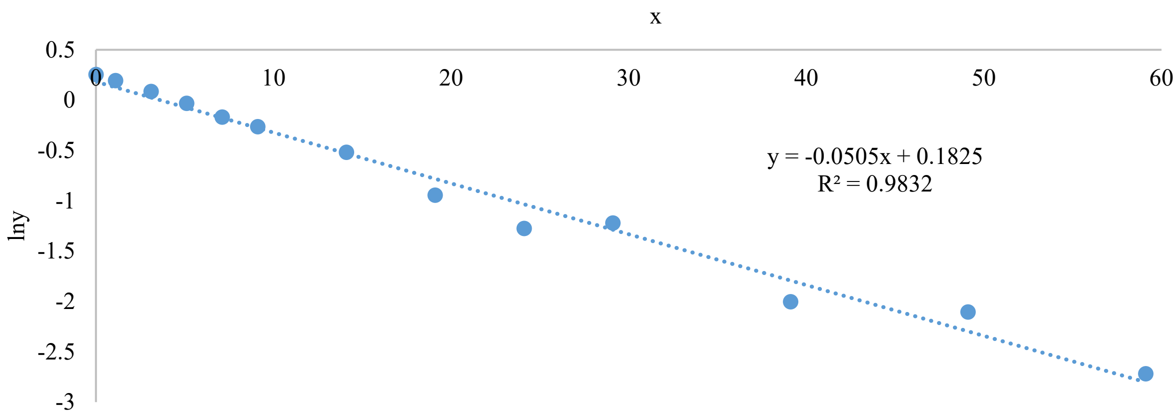

Figure 6. Fit of the linearized exponential model [Eq.(13)].

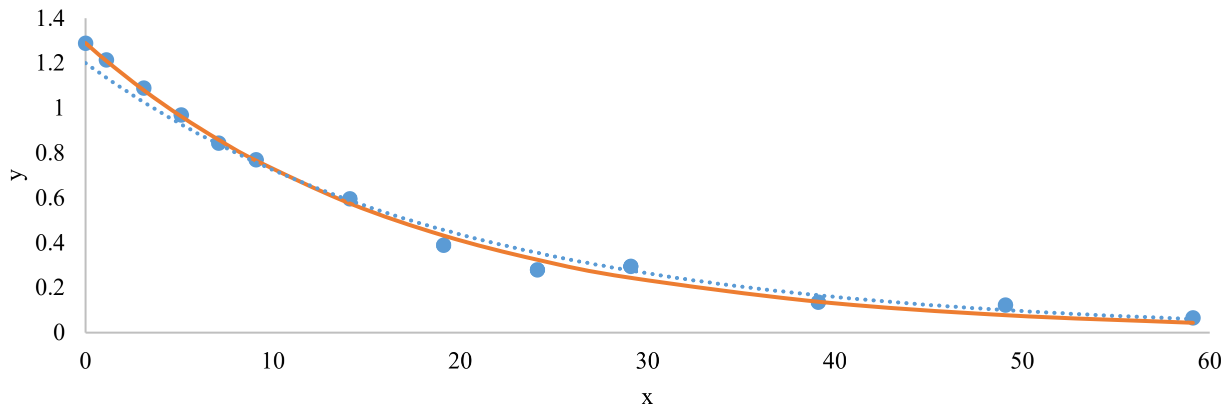

Figure 7. Fit of exponential model [Eq.(12)]. The dotted blue lines were the results of Excel by adding the trend line of exponential model (linear regression). The solid orange lines were the results of Solver Add-In of Excel (nonlinear regression).

We concluded that linear regression fit was worse than that of nonlinear regression. Nevertheless, linearization is not totally useless because parameter values obtained by linear regression can be used as the initial estimates for nonlinear regression (Buzrul, 2024; van Boekel, 2008).

It may be possible to calculate SSE and SST for Excel's exponential model ($1.2002 \cdot e^{-0.0505 \cdot x}$) and the results of these calculations are shown in Table 4.

By using the values in Table 4:

$$R^2 = 1 - \frac{0.03198}{2.2782} = 0.9860 \quad (15)$$

This R^2 value was smaller than the one reported by Excel in Figure 5 which was 0.9926. We can ask the same question once again here: How come Excel could calculate different R^2 than the real R^2? R^2 value reported in Excel for non-linear models was nothing but the square of (Pearson's) correlation coefficient (r). Note that this correlation was not between x and y but between y and y_{model}. Using the y and y_{model} values in Table 4 correlation coefficient (r) was calculated as 0.9953 and its square was 0.9926 which was R^2 value in Figure 5.

Table 4. Sum of squared error (SSE) and sum of squared total (SST) calculations for Excel's exponential model [Eq.(12)] and for the data given in Table 3

x	y	y_{model} ($1.2002 \cdot e^{-0.0505 \cdot x}$)	$(y - y_{model})^2$	$(y - y_{mean})^2$
0.0	1.288	1.2002	0.00771	0.4468
1.1	1.214	1.1353	0.00619	0.3534
3.1	1.088	1.0263	0.00381	0.2195
5.1	0.969	0.9277	0.00171	0.1221
7.1	0.844	0.8386	0.00003	0.0504
9.1	0.769	0.7580	0.00012	0.0223
14.1	0.596	0.5889	0.00005	0.0006
19.1	0.389	0.4575	0.00469	0.0531
24.1	0.279	0.3554	0.00583	0.1160
29.1	0.295	0.2761	0.00036	0.1053
39.1	0.135	0.1666	0.00100	0.2348
49.1	0.122	0.1006	0.00046	0.2475
59.1	0.066	0.0607	0.00003	0.3064
y_{mean} = 0.6195			SSE: 0.03198	SST: 2.2782

It should be noted that both examples (linear model with no-intercept and nonlinear model) could be solved by using standard statistical software package without a problem. We used SigmaPlot 12.0 for both dataset and managed to obtain the correct results (not shown). However, those using Excel for such problems should be aware of these pitfalls.

Conclusion

Use of Excel for regression analysis, and some advantages and disadvantages of Excel for such analysis has been discussed. Some errors for regression analysis with the use of Excel has been also shown. Being aware of those errors are important otherwise, misjudgment such as selecting a wrong model to describe the data or finding the incorrect parameter estimates should not come as a surprise.

Examples shown in this paper revealed that for linear models with no-intercept R^2 and R^2_{adj} values should be recalculated because the incorrect values were displayed by Excel. For nonlinear models in Excel such as exponential model (or power model):

- Excel reports the linear parameter estimates of exponential model (same is also true for the power model).
- Excel calculates R^2 incorrectly and it should be recalculated for the linear parameters.
- One can find the best (real) parameter estimates as well as real R^2 value of the exponential model (or power model) by using nonlinear regression (Solver in Excel).

It is normal to expect new improvements from Microsoft to fix the problems in Excel presented here.

Declarations

Declaration of competing interest

None

Data availability

Data and examples used in this paper are available upon request from the author.

References

Brown, A.M. (2001). A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsheet. *Computer Methods and Programs Biomedicine*, 65, 191–200. https://doi.org/10.1016/S0169-2607(00)00124-3

Buzrul, S. (2024). Fen bilimleri ve mühendislik uygulamalarında deneysel verilerin matematik modellerle tanımlanması. *Excel Uygulamalı Anlatım*. Ankara, Türkiye, *Akademisyen Yayınevi*. http://dx.doi.org/10.37609/akya.3030

De Levie, R. (2012). Advanced Excel for scientific data analysis. New York, NY, USA: *Oxford University Press*.

Denton, P. (2000). Analysis of First-order kinetics using Microsoft Excel Solver. *Journal of Chemical Education*, 77, 1524-1525. https://doi.org/10.1021/ed077p1524

Dolan, K.D., & Mishra, D.K. (2013). Parameter estimation in food science. *Annual Reviews in Food Science and Technology*, 4, 401-422. https://dx.doi.org/10.1146/annurev-food-022811-101247

Halabi, A., Deglaire, A., Hamon, P., Bouhallab, S., Dupont, D., & Croguennec, T. (2020). Kinetics of heat-induced denaturation of proteins in model infant milk formulas as a function of whey protein composition. *Food Chemistry*, 302, 125296. https://doi.org/10.1016/j.foodchem.2019.125296

Harmer, N.J., & Hill, A.M. (2021). Unique data sets and bespoke laboratory videos: teaching and assessing of experimental methods and data analysis in a pandemic. *Journal of Chemical Education*, 98, 4094-4100. https://doi.org/10.1021/acs.jchemed.1c00853

Harris, D.C. (1998). Nonlinear least-squares curve fitting with Microsoft Excel Solver. *Journal of Chemical Education*, 75, 119-121. https://doi.org/10.1021/ed075p119

Keeling, K.B. & Pavur, R.J. (2011). Statistical accuracy of spreadsheet software. The American Statistician 65, 265–273. https://doi.org/10.1198/tas.2011.09076

Kemmer, G., & Keller, S. (2010). Nonlinear least-squares data fitting in Excel spreadsheets. *Nature Protocols*, 5, 267–281. https://doi.org/10.1038/nprot.2009.182

Leylak, C., Yurdakul, M., & Buzrul, S. (2020). Use of Excel in food science 1: linear regression (in Turkish). *Food and Health*, 6, 186-198. https://doi.org/10.3153/FH20020

Mélard, G. (2014). On the accuracy of statistical procedures in Microsoft Excel 2010. *Computational Statistics*, 29, 1095-1128. https://doi.org/10.1007/s00180-014-0482-5

Rubin, S.J., & Abrams, B. (2015). Teaching fundamental skills in Microsoft Excel to first-year students in quantitative Analysis. *Journal of Chemical Education*, 92, 1840-1845. https://doi.org/10.1021/acs.jchemed.5b00122

Serment-Moreno, V. (2021). Microbial modeling needs for the nonthermal processing of foods. *Food Engineering Reviews*, 13, 465-489. https://dx.doi.org/10.1007/s12393-020-09263-8

van Boekel, M.A.J.S. (1996). Statistical aspects of kinetic modeling for food science problems. *Journal of Food Science*, 61, 477-485, 489. https://doi.org/10.1111/j.1365-2621.1996.tb13138.x

van Boekel, M.A.J.S. (2008). Kinetic Modeling of Reactions in Foods. Boca Raton, CRC Press. https://doi.org/10.1201/9781420017410

van Boekel, M.A.J.S. (2022). Kinetics of heat-induced changes in dairy products: Developments in data analysis and modelling techniques. *International Dairy Journal*, 126, 105187. https://doi.org/10.1016/j.idairyj.2021.105187

Yurdakul, M., Leylak, C., & Buzrul, S. (2020). Use of Excel in food science 2: non-linear regression (in Turkish). *Food and Health*, 6, 199-212. https://doi.org/10.3153/FH20021