

Gıda bilimlerinde Excel kullanımı 1: Doğrusal regresyon

Cansu Leylak¹, Merve Yurdakul², Sencer Buzrul¹

Cite this article as:

Leylak, C., Yurdakul, M., Buzrul, S. (2020). Gıda bilimlerinde Excel kullanımı 1: Doğrusal regresyon. *Food and Health*, 6(3), 186-198.

<https://doi.org/10.3153/FH20020>

¹ Konya Gıda ve Tarım Üniversitesi,
Gıda Mühendisliği Bölümü, Melikşah
Mah. Beyşehir Cad. No:9, 42080
Meram, Konya, Türkiye

² Konya Gıda ve Tarım Üniversitesi,
Biyomühendislik Bölümü, Melikşah
Mah. Beyşehir Cad. No: 9, 42080
Meram, Konya, Türkiye

ORCID IDs of the authors:

C.L. 0000-0003-2393-0545

M.Y. 0000-0002-5597-4692

S.B. 0000-0003-2272-3827

Submitted: 14.02.2020

Revision requested: 07.03.2020

Last revision received: 07.03.2020

Accepted: 11.03.2020

Published online: 22.06.2020

Correspondence: Sencer BUZRUL

E-mail: sencer.buzrul@gidatarim.edu.tr



©Copyright 2020 by ScientificWebJournals

Available online at

<http://jfh.sciencwebjournals.com>

ÖZ

Excel genellikle kullandığımız bilgisayarlarda Microsoft® Office'in bir parçası olarak yüklü olarak gelmekte ve deneysel verilerle uğraşan hemen hemen herkes Excel'in basit de olsa kullanımına aşina olmaktadır. Gıda bilimlerinde de deneysel verileri işlemek, dönüştürmek, grafik haline getirmek ya da herhangi bir modelle tanımlamak için Excel'i kullanmak çok yaygındır. Doğrusal regresyon sadece düz bir çizgiyi veriye uydurmak için kullanılır gibi yanlış bir kanı vardır. Ancak, ikinci dereceden bir polinom da ya da bir eğri de doğrusal regresyon kullanılarak veriye uydurulabilir: eğer deneysel verileri tanımlamak için kullanılan model parametresine/parametrelerine göre doğrusalsa bu modelin parametresi/parametreleri doğrusal regresyon kullanılarak bulunabilir. Excel'deki veri çözümleme aracının içerisinde yer alan regresyon uygulaması parametrelerine göre doğrusal modeller için kullanılabilir. Bu çalışmanın amacı doğrusal regresyon kullanılarak parametrelerinin elde edilebileceği modellerin deneysel verilere Excel kullanılarak nasıl uygulanacağını örnekler üzerinde göstermektir. İlk örnekte Excel'in içinde yer alan doğrusal model kullanılarak mikrobiyolojik veriler üzerinde veri çözümleme aracının uygulaması ve sonuçların yorumlanması gösterilmiştir. İkinci örnekte gaz kromatografisi verisini tanımlamak için Excel'in içinde yer almayan ancak kullanıcı tarafından denklemin bilinen bir modelin, üçüncü örnekte ise mikrobiyal büyüme hızını pH'a göre tanımlamak için kullanıcının kendi yarattığı bir modelin uygulamaları gösterilmiştir. Bu çalışmanın gıda mühendisliği ve gıda bilimleri alanında çalışanlar için önemli katkıları olacağı değerlendirilmektedir.

Anahtar Kelimeler: Matematik modeller, Veri Çözümleme, Excel, Doğrusal Regresyon

ABSTRACT

Use of Excel in food science 1: Linear regression

Excel is usually included in the computer package as a part of Microsoft® Office. Almost everyone who deals with the experimental data is more or less familiar with the use of Excel. In food science, it is very common to use Excel to process, transform, sketch or define experimental data with any model. There is a wrong opinion as linear regression can only be used to fit a linear line to the data. However, a second order polynomial or a curved data could also be modelled by using the linear regression: if the model which is used to define the experimental data is linear according to its parameter(s), the parameter(s) of this model can be obtained by using linear regression. Regression application in data analysis tool in Excel could be used for linear models according to its parameters. The aim of this study was to show the application of models to the experimental data by using Excel with examples, where model parameters can be obtained by using linear regression. In the first example, using the linear model in Excel, the application of the data analysis tool on the microbiological data and the interpretation of the results were shown. In the second example, the application of a model that is not included in Excel but its equation is known by the user was shown to define the gas chromatography data. In the third example, the application of a model created by the user to define the microbial growth rate according to pH was shown. It is considered that this study would have important contributions for those working in the field of food engineering and food science.

Keywords: Mathematical models, Data Analysis, Excel, Linear Regression

Giriş

Fen bilimleri ve mühendislik alanlarında, özellikle biyoloji, kimya, biyoteknoloji ve gıda bilimlerinde, deneysel verilerin uygun bir matematik modelle tanımlanması yaygın bir uygulamadır. Bir veya birden çok matematik modelin veriye uydurulması için farklı bilgisayar programları mevcut olup, bu programların kullanımı belli bir deneyim ve uzmanlık gerektirmektedir. Dahası bu programlar (örneğin SigmaPlot ve Microcal Origin gibi programlar) çoğunlukla ücretlidir. Öte yandan kullandığımız bilgisayar ister masaüstü ister dizüstü olsun genel olarak Excel bunların içinde, Microsoft® Office'in bir parçası olarak, yüklü olarak gelmekte ve deneysel verilerle uğraşan hemen hemen herkes Excel'in basitte olsa kullanımına aşina olmaktadır.

Excel hem doğrusal hem de doğrusal olmayan modeller için kullanılabilir. Ancak doğrusal modeller için "Veri Çözümleme" aracının doğrusal olmayan modeller için "Çözücü" aracının kullanılması gerekmektedir. Bu araçların kullanımı her ne kadar çok karmaşık olmasa da belli protokollerin yerine getirilmesini gerektirmektedir. Bu çalışmanın amacı doğrusal modellerin diğer bir deyişle doğrusal regresyon kullanılarak model parametrelerinin elde edilebileceği modellerin deneysel verilere Excel kullanılarak nasıl uygulanacağını göstermektir. Ayrıca model özetinin nasıl yorumlanması gerektiği, modelin veriye uygun olup olmadığının değerlendirilmesi ve grafiksel gösterimi de ele alınacaktır.

Modellerde Doğrusallık

Modeller karmaşık deney verilerinin basit matematiksel ifadelerle tanımlanmasını sağlarlar ve çok farklı şekillerde karşımıza çıkabilmektedirler (logaritmik, üstel veya türevsel denklem). Bunların tamamına kısaca matematik modeller denmektedir. Örneğin $y = f(x)$ denklemini düşünürsek burada x bağımsız değişken olup, araştırmacı tarafından kontrol edilebilir, y bağımlı değişkendir çünkü x 'e bağlıdır ve araştırmacı tarafından ölçülür, f ise fonksiyon olup, veriyi tanımlamak için bir veya birden fazla parametreye sahip olabilir (Brown, 2001). Çok basit bir denklem olan $y = ax + b$ denkleminde a ve b model parametreleri, x ve y de sırasıyla bağımsız ve bağımlı değişkenlerdir. Peki, bu modelin doğrusallığı hakkında ne diyebiliriz? Hemen herkes bu modelin doğrusal olduğunu söyleyecektir. Gerçekten de bu model hem bağımsız değişkenine yani x 'e göre hem de parametrelerine (a ve b) göre doğrusaldır.

Öte yandan $y = Ax^2 + Bx + C$ denkleminde bakarsak parametrelerine göre bu modelin de doğrusal olduğunu görürüz. Ancak, x 'e göre bu model doğrusal değildir. Regresyonda amaç model parametrelerini elde etmek olduğundan ve yukarıda örnek olarak verdiğimiz her iki modelde parametrelerine göre doğrusal olduğundan her iki modele de doğrusal regresyon

uygulanabilir. Bir modelin parametrelerine göre doğrusal olup olmadığını anlamının en kolay yolu modelin ilgili parametreye göre kısmi türevini almaktır (van Boekel ve Zwietering, 2007). Eğer model parametresi kısmi türevde yer almıyorsa model o parametreye göre doğrusaldır.

Örneğin $y = Ax^2 + Bx + C$ modelinin A parametresine göre kısmi türevi $\frac{\partial y}{\partial A} = x^2$, B parametresine göre kısmi türevi $\frac{\partial y}{\partial B} = x$, C parametresine göre kısmi türevi ise $\frac{\partial y}{\partial C} = 1$ 'dir. Görüldüğü gibi her üç parametrede kısmi türevlerinde yer almadığından model parametrelerine göre doğrusaldır ve veriyi bu modelle tanımlamak için doğrusal regresyon kullanılabilir. Genellikle doğrusal regresyon sadece düz bir çizgiyi veriye uydurmak için kullanılır gibi yanlış bir kanı vardır. Oysa görüldüğü gibi ikinci dereceden bir polinom da ya da bir eğri de doğrusal regresyon kullanılarak veriye uydurulabilir.

Excel Kullanarak Deneysel Verilere Doğrusal Regresyon Analizi Yapmak

Bu bölümde basitten karmaşığa üç farklı örnek üzerinde Excel'de "Veri Çözümleme" aracını kullanarak doğrusal regresyon ile deneysel verileri tanımlamayı göstereceğiz. Ancak, veri çözümleme aracı Excel'de yüklü değilse (Microsoft Office Standard 2016 Excel'de "Veri" sekmesine tıkladığımızda sağ üste veri çözümleme görünmüyorsa), sırasıyla Dosya > Seçenekler > Eklentiler > Excel Eklentileri (Git) sekmelerinden sonra çıkan ekrandan "Çözümleme Araç Takımı" ve "Çözümleme Araç Takımı - VBA" işaretleyerek yüklenmelidir.

Listeria monocytogenes'in İnaktivasyonunun Doğrusal Modelle Tanımlanması

İlk örneğimizde sıvı besiyerinde ısıyla inaktive edilen ve hastalık yapıcı bir bakteri olan *L. monocytogenes*'i ele alacağız. *L. monocytogenes* pastörizasyon sıcaklıklarına (71.7°C'de 15 saniye veya 62.8°C'de 30 dakika) karşı çok hassas olduğundan (Ray, 2014) bu bakterinin daha düşük bir sıcaklıktaki (58°C'de) inaktivasyonu örnek olarak kullanılmıştır. Şekil 1'de *L. monocytogenes*'in mL'de ölçülen canlı hücre sayısı (\log_{10} KOB/mL, KOB: koloni oluşturan birim) zamana bağlı olarak gösterilmektedir. Örneğimizdeki bağımsız değişken zaman (dakika) iken bağımlı değişken canlı hücre sayısıdır (\log_{10} KOB/mL).

Zaman ve zamana karşı ölçülen canlı hücre sayısı x ve y olacak şekilde bir Excel sayfasında tanımlanır (Şekil 1). Daha sonrasında verilerin regresyon uygunluğunu saptamak amacıyla dağılım grafiği çizilir. Doğrusal model kullanarak bu veriyi tanımlamak mümkündür:

$$\log_{10} N(t) = \log_{10} N_0 - k \cdot t \quad (1)$$

Burada $\log_{10} N(t)$ bakterinin t zamandaki sayısı, $\log_{10} N_0$ bakterinin başlangıçtaki sayısı ($t = 0$), k ise inaktivasyon hızıdır (zaman⁻¹). Yani doğrusal modelin iki parametresi $\log_{10} N_0$ ve k 'dir.

Tablo 1. Örnek 3'te yer alan modellerin uyumlarının karşılaştırılması

Table 1. Comparison of goodness-of-fit of the models in Example 3

Model	Ayarlı R ²	Standart hata
2'nci dereceden polinom	0.9031	0.0430
3'üncü dereceden polinom	0.9411	0.0335
Alternatif model	0.9399	0.0338

Grafik üzerindeki verilere sağ tıklanarak "Eğilim çizgisi ekle"den Excel'in içindeki "Doğrusal model" seçilir ve en altta yer alan "Grafik üzerinde Denklemi görüntüle" ile "Grafik üzerinde R-kare değerini görüntüle" kısımları işaretlenir. Grafik üzerinde model uyumu (kesikli mavi çizgiler), model denklemi ($y = -0.9794x + 10.287$) ve R-kare değeri ya da belirleme katsayısı ($R^2 = 0.9932$) görünmektedir (Şekil 2). Bu durumda, Denklem 1 ile grafik üzerinde görünen denklemi eşleştirirsek *L. monocytogenes*'in başlangıçtaki ($t = 0$) sayısını $\log_{10} N_0 = 10,287 \log_{10} \text{KOB/mL}$ ve inaktivasyon hızını $k = 0.9794 \text{ dakika}^{-1}$ olarak bulabiliriz.

Burada birçok Excel kullanıcısı model parametrelerini ($\log_{10} N_0$ ve k) ve R^2 değerini elde etmekle istediğimiz sonuca ulaştığımızı düşüncektir. Ancak, mevcut durum istediğimiz sonucu elde edebilmekten çok uzaktadır. Birincisi model parametreleri belirsizlikleri (standart hata veya güven aralığı) ile elde edilmemiştir ve bu halleriyle hiçbir anlam taşımamaktadırlar (van Boekel, 1996; 2008). İkincisi her ne kadar birçok bilimsel çalışmada hala model uyumunu belirlemede kullanılsa da R^2 değeri model uyumunu anlamak için yeterli değildir (Ratkowsky, 2004). Daha doğrusu R^2 değerinin 1'e yakın olması ($R^2 > 0.90$) model uyumunun iyi olduğunu göstermez (Montgomery ve Runger, 2011) ancak R^2 değeri düşükse ($R^2 < 0.70$) model uyumuyla ilgili bir sorun olduğu aşıkardır (Granato ve ark., 2014). Herhangi bir model uyumu için yüksek R^2 değeri elde edilmesi durumunda o modelin uyumunu değerlendirmek için başka uyum ölçütlerini de gözden geçirmek en doğru yoldur (Granato ve ark., 2014).

İşte tam da bu noktada "Veri Çözümleme" aracını kullanmak istediğimiz sonuçlara ulaşmayı mümkün kılacaktır. Bunun için Excel'de veri çözümleme aracı açılır ve çıkan pencereden regresyon seçilir (Şekil 3a). Regresyon altında Y giriş aralığı olarak *Listeria monocytogenes*'in canlı hücre sayısı

(Excel hücreleri B2:B13) ve X giriş aralığı süre (Excel hücreleri A2:BA13) olarak seçilir (Şekil 3b). Veriler uygun şekilde girildikten sonra başka hiçbir değişiklik yapmadan Tamam'a basılır ve yeni bir Excel çalışma sayfasında Şekil 4'te görünen sonuçlar elde edilir. Görüldüğü gibi uygulanan doğrusal regresyon sonucu birçok bilgi ekranda belirmiştir.

Sonuçları değerlendirmeye "Özet Çıkışı"nın en altındaki tablodan başlarsak, grafik gösteriminde elde edilen parametrelerin bu sefer standart hataları ($\log_{10} N_0 = 10.29 \pm 0.08$ ve $k = 0.98 \pm 0.026$) ve % 95 güven aralıkları ($\log_{10} N_0 = 10.29 \pm 0.19$ ve $k = 0.98 \pm 0.06$) ile birlikte bulunduğu görülebilir. Parametre belirsizliklerini elde etmek özellikle gıda bilimlerinde önemlidir (Dolan ve Mishra, 2013) ve % 95 güven aralığı parametre değerinin % 95 ihtimalle olacağı aralığı belirtmektedir. Sonuçlarda görülen p -değerinin küçük olması elde edilen parametrelerin istatistiksel olarak anlamlı olduğunu göstermektedir. Genel olarak $p > 0.05$ ise parametreler istatistiksel olarak anlamsızdır ve verilerimizi tanımlamak için anlamsız parametrelere sahip modeli kullanmak geçerli bir uygulama değildir. Örneğimizde $\log_{10} N_0$ parametresinin p -değeri 3×10^{-17} , k parametresinin p -değeri ise 3.7×10^{-12} 'dir. Her iki parametrenin de p -değeri < 0.05 olduğundan her iki parametrede anlamlıdır. Parametrelerin standart hata ve güven aralığı da p -değeri ile orantılıdır. Diğer bir deyişle p -değeri ne kadar küçükse standart hata/güven aralığı o kadar küçüktür. Aynı tabloda yer alan t -değeri ise parametre değerinin standart hataya bölümüdür ve sayısal olarak büyük olması parametrenin anlamlı olduğunu gösterir.

Özet çıkışının ortasında ANOVA tablosu yer almaktadır (Şekil 4). Burada df (degrees of freedom) serbestlik derecesi demektir. Toplam 12 verimiz olduğundan (12 zamanda elde edilen 12 canlı hücre sayısı) toplamın serbestlik derecesi $12-1 = 11$ 'dir. Veri sayısı özet çıkışının ilk tablosunun hemen altında "Gözlem" olarak da gösterilmektedir (Şekil 4). Regresyonda veya modelde iki parametremiz ($\log_{10} N_0$ ve k) olduğundan regresyonun serbestlik derecesi $2-1 = 1$ 'dir. Farkın serbestlik derecesi ise bu ikisinin farkıdır yani $11 - 1 = 10$ 'dur. Farkın serbestlik derecesini veri sayısını (12) modelin parametre sayısından (2) çıkartarak da hesaplamak mümkündür ($12-2 = 10$). Tabloda görülen SS (Sum of squares) karelerin toplamı anlamını taşımaktadır. Örneğimiz için toplamın SS 'si aşağıda verilmiştir:

$$SS_{\text{Toplam}} = \sum_{n=1}^{12} (y_{\text{veri}} - y_{\text{veri,ortalama}})^2 \quad (2)$$

Burada her bir y verisini y verilerinin ortalamasından ($y_{\text{veri,ortalama}} = 7,59$) çıkartarak karelerini alıp bütün bunları toplarsak $SS_{\text{Toplam}} = 34,53$ olarak hesaplanabilir. Regresyonun (modelin) SS 'si aşağıdaki gibi hesaplanabilir:

$$SS_{Model} = \sum_{n=1}^{12} (y_{model} - y_{veri,ortalama})^2 \quad (3)$$

Model denklemleri ve model parametreleri kullanılarak her bir zaman için y_{model} hesaplanırsa buradan $SS_{model} = 34.29$ olarak bulunabilir. SS_{Toplam} veri noktalarının ortalama değerden sapmasını ölçerken SS_{Model} ise modelden elde edilen veri noktalarının ortalama değerden sapmasını ölçer. Farkın SS 'si $SS_{Toplam} - SS_{Model} = 0.24$ olarak elde edilir. Öte yandan farkın SS 'sini aşağıdaki gibi hesaplamak ta mümkündür:

$$SS_{Fark} = \sum_{n=1}^{12} (y_{veri} - y_{model})^2 \quad (4)$$

Regresyonda amaç farkların karelerinin (SS_{Fark}) toplamının en aza indirilmesidir. Dolayısıyla SS_{Fark} 'ın mümkün olan en düşük değerde olması istenir. Denklem 2, 3 ve 4 arasındaki bağıntıyı bulmak çok da zor değildir:

$$SS_{Toplam} = SS_{Model} + SS_{Fark} \quad (5)$$

Tabloda görülen MS (Mean square) ortalama kare değeri anlamına gelir SS değerinin serbestlik derecesine bölümü ile elde edilir yani:

$$MS_{Model} = \frac{SS_{Model}}{df_{Model}} = \frac{34,29}{1} = 34,29 \quad (6)$$

$$MS_{Fark} = \frac{SS_{Fark}}{df_{Fark}} = \frac{0,24}{10} = 0,024 \quad (7)$$

F -değeri bu ikisinin oranıdır: $MS_{Model}/MS_{Fark} = 34.29/0.024 = 1451.27$. Eğer F -değeri örneğimizdeki gibi yüksek bir sayı ise bağımsız değişkenin (x ya da örneğimizde zaman) bağımlı değişkeni (y ya da örneğimizde canlı hücre sayısı) tanımlamak için etkisi olduğu söylenebilir. Öte yandan "anlamlılık F -değeri" ne kadar düşük ise (örneğimizde bu değer 3.7×10^{-12} 'dir) bağımsız değişkenin bağımlı değişkeni tanımlamak için kullanılabileceği anlamı çıkar.

Özet çıkışının en üstündeki tabloda modelin uyum iyiliğini gösteren ölçümler görülmektedir. Grafik gösterimi sadece R^2 değerini verirken burada R^2 'ye ilaveten ayarlı R^2 (adjusted R^2) ve modelin standart hatasını da görmekteyiz. R^2 değeri bir altta yer alan ANOVA tablosundaki değerler kullanılarak hesaplanabilir:

$$R^2 = 1 - \frac{SS_{Fark}}{SS_{Toplam}} = 1 - \frac{0,24}{34,53} = 0,9932 \quad (8)$$

Ayarlı R^2 ise aşağıdaki denklemden hesaplanabilir ve 0.9925 olarak bulunabilir.

$$R_{ayarlı}^2 = 1 - (1 - R^2) \cdot \frac{df_{Toplam}}{df_{Fark}} \quad (9)$$

Ayarlı R^2 sadece iki durumda R^2 'ye eşit olabilir: (i) modelin tek bir parametresi varsa yani $df_{Toplam} = df_{Fark}$ ise; (ii) model mükemmel uyumlu ise yani $R^2 = 1$ ise. Onun dışında ayarlı R^2 her zaman R^2 'den düşük çıkacaktır. Bazı kaynaklarda R^2 yerine ayarlı R^2 değerinin kullanılmasının daha uygun olduğu belirtilmekle birlikte (Davey ve Amos, 2002) model uyumunun en uygun göstergesi modelin standart hata değeridir (Ratkowsky, 2004). Standart hata (SH) aşağıdaki gibi hesaplanır:

$$SH = \sqrt{MS_{Fark}} = \sqrt{0,024} = 0,1537 \quad (10)$$

Standart hataya ortalama karesel hataların karekökü de (RMSE: root mean square error) denir. R^2 model tarafından açıklanan bağımlı değişken varyansının yüzdesini temsil ederken standart hata veri noktalarının regresyon (model) çizgisinden düştüğü ortalama mesafeyi gösteren mutlak bir ölçüdür. Diğer deyişle, daha küçük standart hata değerleri daha iyidir çünkü gözlemlerin yerleştirilen hatta (model çizgisine) daha yakın olduğunu gösterir. Standart hatanın birimi y 'nin birimi ile aynıdır. Örneğimizde modelin 0.1537 \log_{10} KOB/mL'lik bir standart hatası vardır ve mikrobiyolojide 0.5 \log_{10} KOB/mL'den az hatalar makul olarak değerlendirildiğinden (Jarvis, 1989; Mossel ve ark., 1995) modelimiz bu veriyi tanımlamak için son derece uygundur.

Kısaca belirtmek gerekirse Excel'de veri çözümleme aracını kullanmayı ve elde edilen sonuçları (özet çıkışını) yorumlamayı bilirsek model parametrelerinin istatistiksel olarak anlamlı olup olmadığını ve modelin veriyi tanımlamak için yeterli olup olmadığını anlamak çok kolaylaşır. Dahası veri çözümleme bizi birçok farklı hesaptan kurtararak (SS hesaplamaları, R^2 , ayarlı R^2 ve standart hata değerleri) zamandan tasarruf sağlamaktadır.

Gaz Kromatografi Verisinin Van Deemter Modeli İle Tanımlanması

İkinci örneğimizde Şekil 5'te görünen verilerin [bağımsız değişken akış hızı (mL/dakika) iken bağımlı değişken tabaka yüksekliğidir (mm)] van Deemter modeli ile tanımlanmasını ele alacağız. Akış hızına karşı tabaka yüksekliği x ve y olacak şekilde yeni bir Excel sayfasında tanımlanan verilerin dağılım grafiği çizilir (Şekil 6). İlk örneğimizde Excel'in içindeki bir denklemi (doğrusal model) kullanma imkanı varken bu örneğimizde bunu yapmak mümkün değildir çünkü Excel'in içerisindeki hiçbir model bu veriyi tanımlayamamaktadır.

Kullanacağımız van Deemter modeli aşağıdaki şekildedir:

$$y = Ax + B/x + C \quad (11)$$

Burada y tabaka yüksekliği (mm), x akış hızı (mL/dakika), A , B ve C ise model parametreleridir. Her şeyden önce model parametrelerine göre doğrusaldır ve bu nedenle bu parametreler doğrusal regresyonla elde edilebilir. Bu da Excel’de veri çözümlenme aracını kullanmak anlamına gelmektedir. Denklem 11’de hem x hem de $1/x$ yer aldığından hemen x sütununun yanına yeni bir sütun eklenerek $1/x$ değerleri Excel’de hesaplatılır (Şekil 7). Veri çözümlenmeden regresyona girildikten sonra Y çıkış aralığı için C2:C14 seçilir. Denklemimizde iki x değeri yer aldığından (x ve $1/x$) X çıkış aralığı için bunların her ikisi de seçilir (A2:A14 + B2:B14) ve özet çıkışı elde edilir (Şekil 8). Özet çıkışından model parametrelerinin anlamlı ve modelin verilerle uyumlu olduğu (standart hata = 0,11) anlaşılmaktadır. Ancak, grafik üzerinde görsel olarak da bu uyumluluğu göstermek uygun olacaktır. Bunun için D kolonuna model yazılır ve her bir akış hızı için plaka yüksekliği Excel’e hesaplatılır. Daha sonra grafik üzerindeki verilere sağ tıklayıp “Veri Seç”ten “Ekle”ye gelinir. Çıkan pencereye seri adı olarak “model” yazılır. Seri X değeri olarak x seçilir (A2:A14), seri Y değeri olarak ise hesaplatılan model verileri (D2:D14) girilir. Böylece model verileri (turuncu olarak görülen daireler) deney verileri (daha önce grafik üzerinde gördüğümüz mavi daireler) ile birlikte gözlemlenebilir, ancak bu göz yanılmasına yol açacağından model verilerini görsellik olarak daha uygun hale getirmek yerinde olacaktır. Bunun için model verilerine (turuncu daireler) sağ tıklayıp “Seri Grafik Türünü Değiştir”e gelip açılan pencereden model için dağılım yerine “Düz Çizgilerle Dağılım” seçilir. Böylece deney verileri ve bu verileri tanımlamak için kullanılan modelin uyumu aynı grafik üzerinde görülebilir (Şekil 9).

Mikrobiyal Büyüme Hızının pH’a Göre Tanımlanması

Son örneğimizde Şekil 10’da görünen bir bakterinin pH’ya (x) karşılık büyüme hızını (y) gösteren veriler vardır. Diğer iki örneğimizde olduğu gibi grafik çizerek başlamak en doğru yaklaşım olacaktır. Verilere dikkatlice bakınca Excel’in içinde yer alan 2’nci ve 3’üncü dereceden polinom modellerle verinin tanımlanabileceği anlaşılmaktadır (Şekil 11). Ancak, her iki model de veriyi tanımlama açısından sorunludur. Şöyle ki, 2’nci dereceden polinom modelde büyüme hızı (μ) $6 < \text{pH} < 7$ aralığında (ki bu aralıkta herhangi bir deneysel veri bulunmamaktadır) önce artıp sonra azalmaktadır. Büyüme hızı yaklaşık olarak $\text{pH} = 6.5$ ’te de maksimuma ulaşmaktadır. Diğer yandan aynı aralıkta 3’üncü dereceden polinom modelde büyüme hızı hafifçe azalır sonra artmaktadır ve yine yaklaşık olarak $\text{pH} = 6.5$ ’te bu aralığın en düşük büyüme hızı gözlemlenmektedir (Şekil 11). Veriye baktığımızda genel

olarak pH artıkça büyüme hızının da arttığı görülmektedir dahası mikrobiyoloji bilgimiz genel olarak bakteriler için $\text{pH} = 7$ ’de asidik pH ’lara göre büyüme hızının daha fazla olması gerektiğini söylemektedir. Bu durumda optimum pH ’yı 7 olarak kabul edip yeni bir değişken olarak $\text{pH}' = (7 - \text{pH})^2$ tanımlanırsa ve bu değişken baz alınarak parabolik regresyon yapılırsa [$\mu = a(\text{pH}')^2 + b$] daha önce uygulanan iki polinom modele alternatif olabilir. Burada hem 2’nci hem de 3’üncü dereceden polinom modeller parametrelerine göre doğrusaldır. Dahası alternatif olarak yazmış olduğumuz model de parametrelerine göre doğrusal olduğundan Excel’de veri çözümlenme aracını kullanmak mümkündür. Burada dikkat edilecek husus 2’nci dereceden polinom model için x^2 ’yi Excel’de hesaplatmak ve veri çözümlenme-regresyon’da X giriş aralığına hem x ’li hem de x^2 ’li sütunları girmek; 3’üncü dereceden polinom modelde için ise x^2 ve x^3 ’ü Excel’de hesaplatarak X giriş aralığına x ’li, x^2 ’li ve x^3 ’lü sütunları girmektir. Şekil 12’de her üç modelin veriye uyumlulukları gösterilmektedir. Tablo 1’de de özet çıkışlarından elde edilen ayarlı R^2 ve standart hatalar verilmiştir. İkinci dereceden polinom model en kötü uyumu göstermektedir. Öte yandan, her ne kadar 3’üncü dereceden polinom model en iyi uyumu vermiş gibi görünse de (en yüksek ayarlı R^2 ve en düşük standart hata değerine sahiptir) alternatif model mikrobiyolojik olarak daha gerçekçi sonuç vermektedir. Bunun nedeni yukarıda da değindiğimiz gibi bakterilerin genellikle $\text{pH} = 7$ değerinde büyüme hızlarının maksimum olmasıdır. Alternatif model bize bunu sağlamaktadır (Şekil 12). Oysa 2’nci polinom modelde büyüme hızı yaklaşık 6.5 pH değerinde maksimumken, 3’üncü dereceden polinom modelde büyüme hızı $\text{pH} = 6.5$ ’de azalmakta sonra tekrar artmaktadır. Burada “istatistiksel olarak daha iyi sonuç veren ancak daha fazla parametre içeren bir model mi (4 parametreye sahip 3’üncü dereceden polinom model) yoksa istatistiksel olarak biraz daha kötü ancak daha az parametreye sahip aynı zamanda mikrobiyolojik olarak gerçekçi bir model mi (2 parametreye sahip alternatif model)” sorusunun cevabını da vermiş bulunuyoruz. Alternatif model ile 3’üncü dereceden polinom model arasındaki uyumun farkının az olması (Tablo 1) ve alternatif modelin daha az parametreyle gerçekçi sonuçlar elde etmesi bu modeli tercih sebebi haline getirmektedir.

Sonuç

Eğer deneysel verileri tanımlamak için kullanılan herhangi bir model parametresine/ parametrelerine göre doğrusalsa bu modelin parametresi/parametreleri doğrusal regresyon kullanılarak bulunabilir. Excel’de yer alan veri çözümlenme uygulaması parametrelerine göre doğrusal modeller için kullanılacak güçlü bir doğrusal regresyon aracıdır. Bu çalışmada basitten karmaşığa üç örnekle Excel’in doğrusal regresyonda nasıl kullanılacağı açıklanmaya çalışılmıştır. İlk örnekte hem

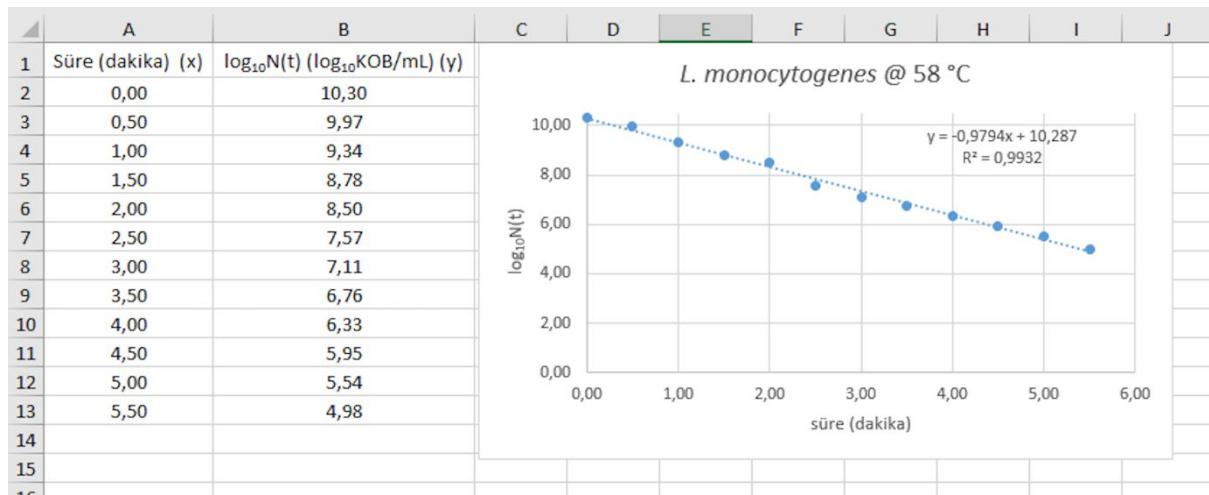
verilerin hem de model parametrelerin doğrusal olduğu ve Excel'in içinde yer alan bir modelin kullanılarak mikrobiyolojik veriler üzerinde veri çözümüleme aracının uygulaması ve sonuçların yorumlanması gösterilmiştir. İkinci örnekte gaz kromatografi verisini tanımlamak için Excel'in içinde yer al-

mayan ancak bilinen bir modelin, üçüncü örnekte ise mikrobiyal büyüme hızını pH'a göre tanımlamak için kullanıcının kendi yarattığı bir modelin uygulamaları gösterilmiştir. Bu çalışmanın gıda mühendisliği ve gıda bilimleri alanında çalışanlar için önemli katkıları olacağı değerlendirilmektedir.

	A	B
1	Süre (dakika) (x)	$\log_{10}N(t)$ (\log_{10} KOB/mL) (y)
2	0,00	10,30
3	0,50	9,97
4	1,00	9,34
5	1,50	8,78
6	2,00	8,50
7	2,50	7,57
8	3,00	7,11
9	3,50	6,76
10	4,00	6,33
11	4,50	5,95
12	5,00	5,54
13	5,50	4,98
14		
15		
16		

Şekil 1. Sıvı besiyerindeki *Listeria monocytogenes*'in 58 °C'deki inaktivasyon verilerinin Excel'e girilmiş hali. Orijinal veriler Hassani ve ark. (2005)'dan alınmıştır.

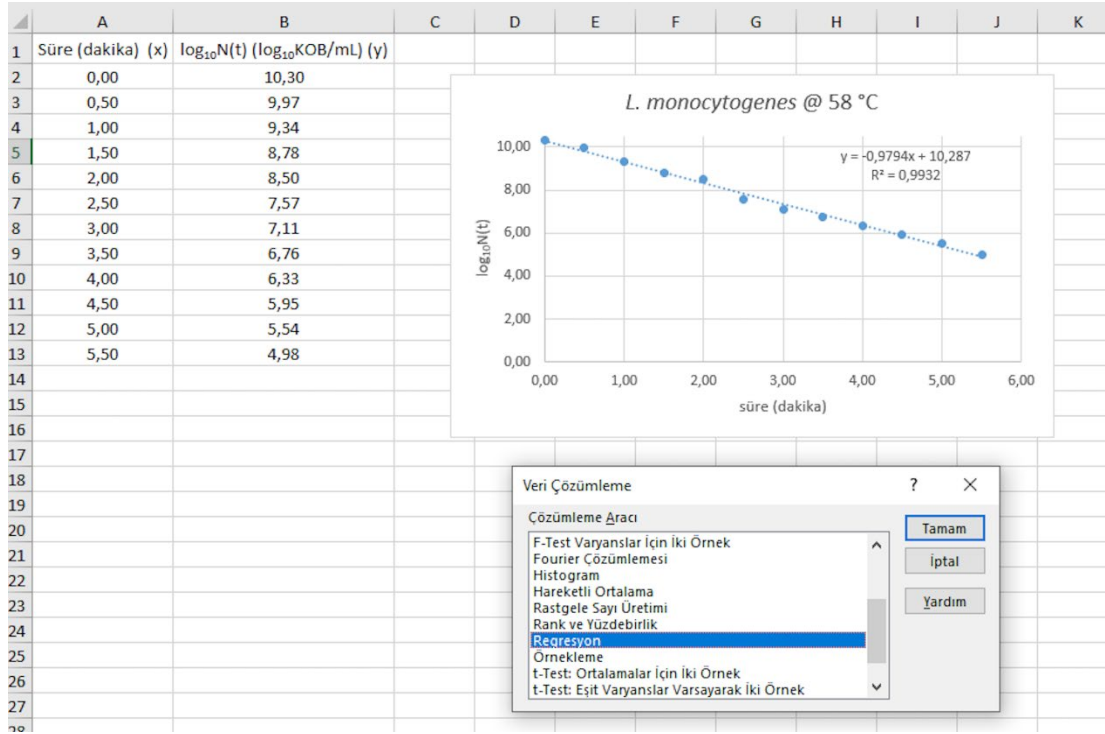
Figure 1. Inactivation data of *Listeria monocytogenes* in broth at 58 °C inserted in Excel. Original data were from Hassani et al. (2005).



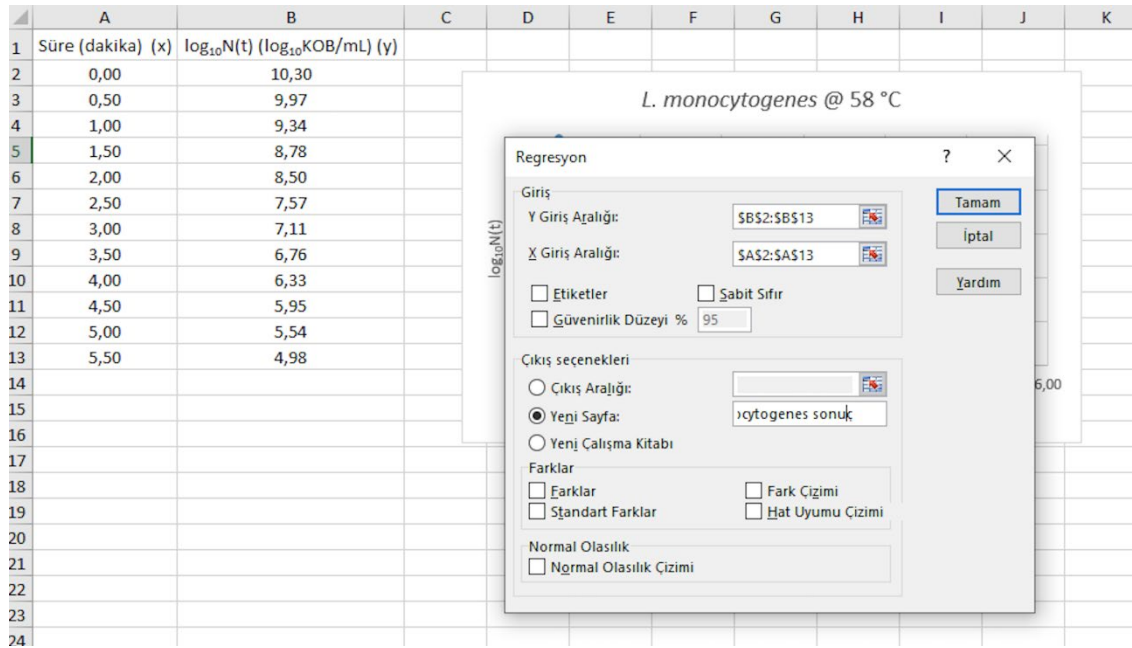
Şekil 2. Şekil 1'de gösterilen verinin Excel'de grafikleştirilmiş hali ve Excel'in içindeki doğrusal modelin ($y = ax + b$) bu veriye uygulaması.

Figure 2. Sketching the graph of the data in Excel given in Figure 1 and application of the linear model ($y = ax + b$) in Excel to the data.

(a)



(b)



Şekil 3. Excel’de Şekil 1’de gösterilen veriye “Veri Çözümleme” aracını kullanarak doğrusal regresyon uygulanması.

Figure 3. Application of linear regression to the data given in Figure 1 by using “Data Analysis” tool in Excel.

	A	B	C	D	E	F	G	H	I
1	ÖZET ÇIKIŞI								
2									
3	<i>Regresyon İstatistikleri</i>								
4	Çoklu R	0,996572442							
5	R Kare	0,993156631							
6	Ayarlı R Kare	0,992472295							
7	Standart Hata	0,153713209							
8	Gözlem	12							
9									
10	ANOVA								
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Anlamlılık F</i>			
12	Regresyon	1	34,29021416	34,29021416	1451,26867	3,70418E-12			
13	Fark	10	0,236277506	0,023627751					
14	Toplam	11	34,52649167						
15									
16		<i>Katsayılar</i>	<i>Standart Hata</i>	<i>t Stat</i>	<i>P-değeri</i>	<i>Düşük %95</i>	<i>Yüksek %95</i>	<i>Düşük 95,0%</i>	<i>Yüksek 95,0%</i>
17	Kesişim	10,2874359	0,083469499	123,2478456	3,03387E-17	10,10145426	10,47341753	10,10145426	10,47341753
18	X Değişkeni 1	-0,979370629	0,025708289	-38,09552034	3,70418E-12	-1,036652266	-0,922088993	-1,036652266	-0,922088993
19									
20									
21									

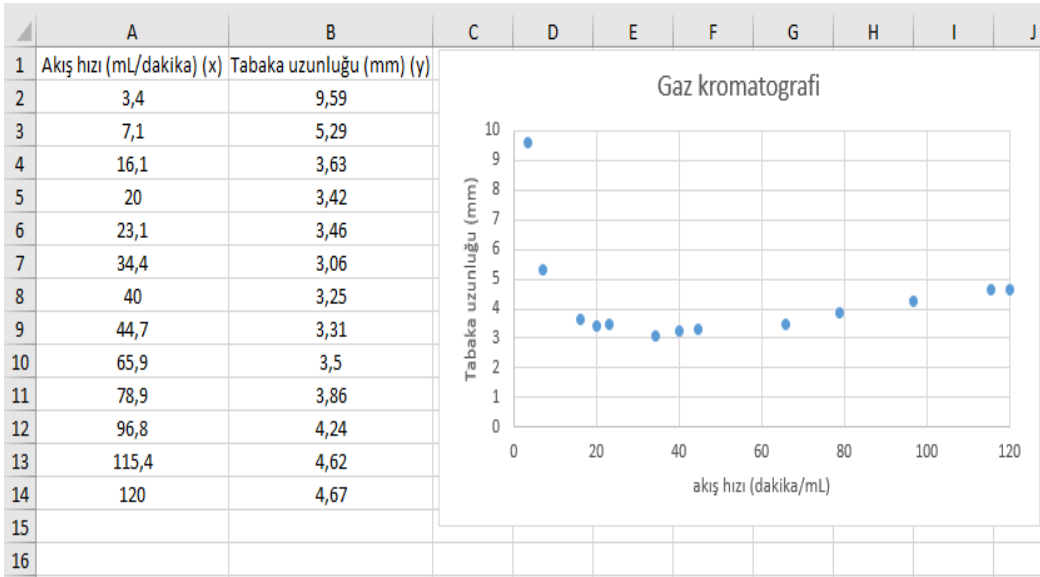
Şekil 4. Şekil 1’de gösterilen veriye uygulanan doğrusal regresyonun Excel’deki “Özet Çıkışı”.

Figure 4. “Summary Output” of the application of linear regression to the data given in Figure 1 in Excel.

	A	B
1	Akış hızı (mL/dakika) (x)	Tabaka uzunluğu (mm) (y)
2	3,4	9,59
3	7,1	5,29
4	16,1	3,63
5	20	3,42
6	23,1	3,46
7	34,4	3,06
8	40	3,25
9	44,7	3,31
10	65,9	3,5
11	78,9	3,86
12	96,8	4,24
13	115,4	4,62
14	120	4,67
15		
16		

Şekil 5. Gaz kromatografi verileri. Orijinal veriler Moody (1982)’den alınmıştır.

Figure 5. Gas chromatography data. Original data were from Moody (1982).



Şekil 6. Şekil 5'te gösterilen verinin Excel'de grafikleştirilmiş hali.

Figure 6. Sketching the graph of the data in Excel given in Figure 5.

	A	B	C
1	Akış hızı (mL/dakika) (x)	(1/x)	Tabaka uzunluğu (mm) (y)
2	3,40	0,29	9,59
3	7,10	0,14	5,29
4	16,10	0,06	3,63
5	20,00	0,05	3,42
6	23,10	0,04	3,46
7	34,40	0,03	3,06
8	40,00	0,03	3,25
9	44,70	0,02	3,31
10	65,90	0,02	3,50
11	78,90	0,01	3,86
12	96,80	0,01	4,24
13	115,40	0,01	4,62
14	120,00	0,01	4,67
15			
16			

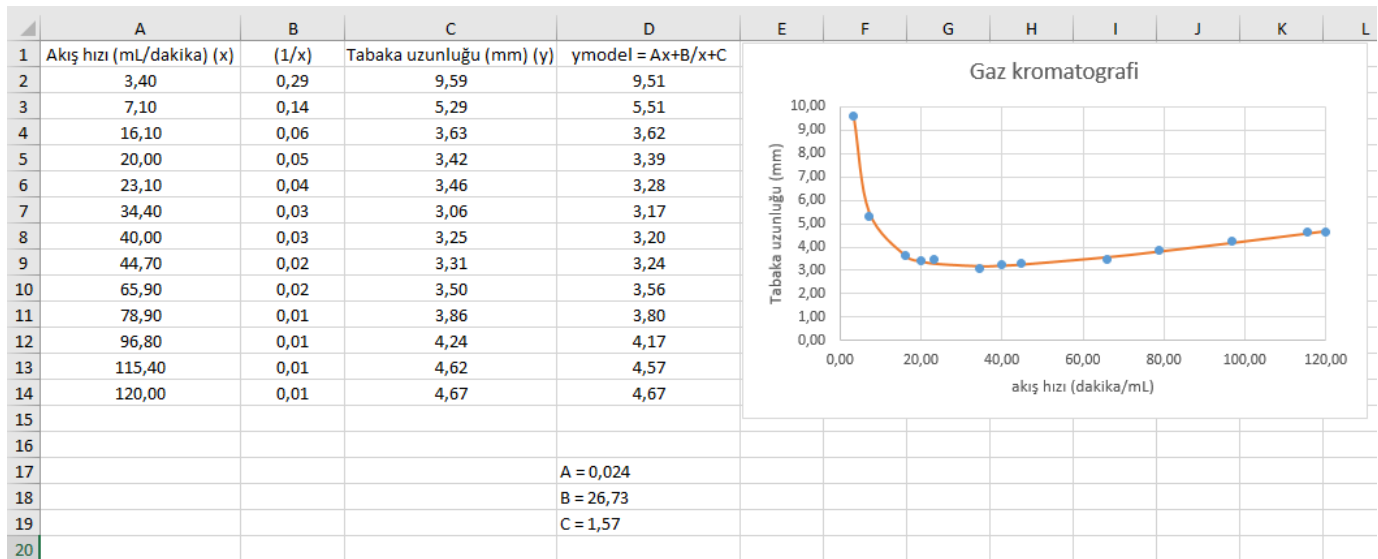
Şekil 7. Şekil 5'te gösterilen x verisinin Excel'de $1/x$ 'e dönüştürülmüş hali.

Figure 7. Transformation of the x data given in Figure 5 to $1/x$ in Excel.

	A	B	C	D	E	F	G	H	I
1	ÖZET ÇIKIŞI								
2									
3	Regresyon İstatistikleri								
4	Çoklu R	0,998405078							
5	R Kare	0,9968127							
6	Ayarlı R Kare	0,99617524							
7	Standart Hata	0,106506046							
8	Gözlem	13							
9									
10	ANOVA								
11		df	SS	MS	F	Anlamlılık F			
12	Regresyon	2	35,47636462	17,73818231	1563,725771	3,28939E-13			
13	Fark	10	0,113435378	0,011343538					
14	Toplam	12	35,5898						
15									
16		Katsayılar	Standart Hata	t Stat	P-değeri	Düşük %95	Yüksek %95	Düşük 95,0%	Yüksek 95,0%
17	Kesişim	1,568074804	0,075571995	20,74941642	1,49738E-09	1,399689904	1,736459703	1,399689904	1,736459703
18	X Değişkeni 1	0,02435858	0,000966832	25,19423027	2,22317E-10	0,022204345	0,026512816	0,022204345	0,026512816
19	X Değişkeni 2	26,72786047	0,487411668	54,8363164	9,85846E-14	25,64183959	27,81388134	25,64183959	27,81388134
20									
21									
22									
23									

Şekil 8. Şekil 7’de gösterilen veriye uygulanan doğrusal regresyonun [van Deemter modeli ($y = Ax + B/x + C$)] Excel’deki “Özet Çıktısı”.

Figure 8. “Summary Output” of the application of linear regression [van Deemter model ($y = Ax + B/x + C$)] to the data given in Figure 7 in Excel.



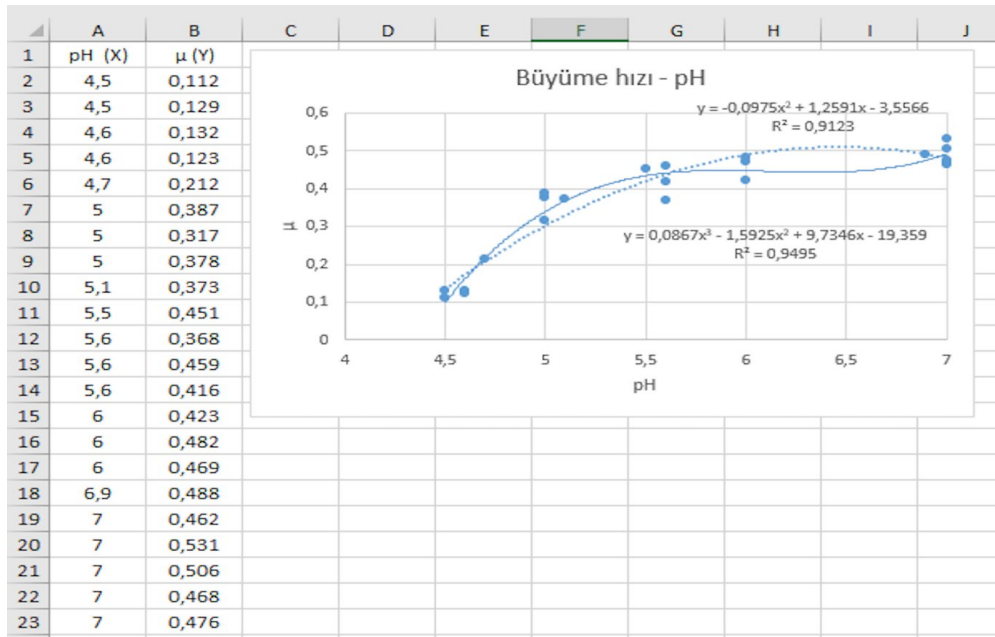
Şekil 9. Şekil 7’de gösterilen veriye uygulanan van Deemter modelinin ($y = Ax + B/x + C$) uyumunun Excel’de grafik üzerinde gösterilmesi. Mavi daireler deney verilerini turuncu çizgi ise model uyumunu belirtmektedir.

Figure 9. Showing the application of the van Deemter model ($y = Ax + B/x + C$) on the graph to the data given in Figure 7 in Excel. Blue circles represent experimental data orange line represents model fit.

	A	B
1	pH (X)	μ (Y)
2	4,5	0,112
3	4,5	0,129
4	4,6	0,132
5	4,6	0,123
6	4,7	0,212
7	5	0,387
8	5	0,317
9	5	0,378
10	5,1	0,373
11	5,5	0,451
12	5,6	0,368
13	5,6	0,459
14	5,6	0,416
15	6	0,423
16	6	0,482
17	6	0,469
18	6,9	0,488
19	7	0,462
20	7	0,531
21	7	0,506
22	7	0,468
23	7	0,476
24		

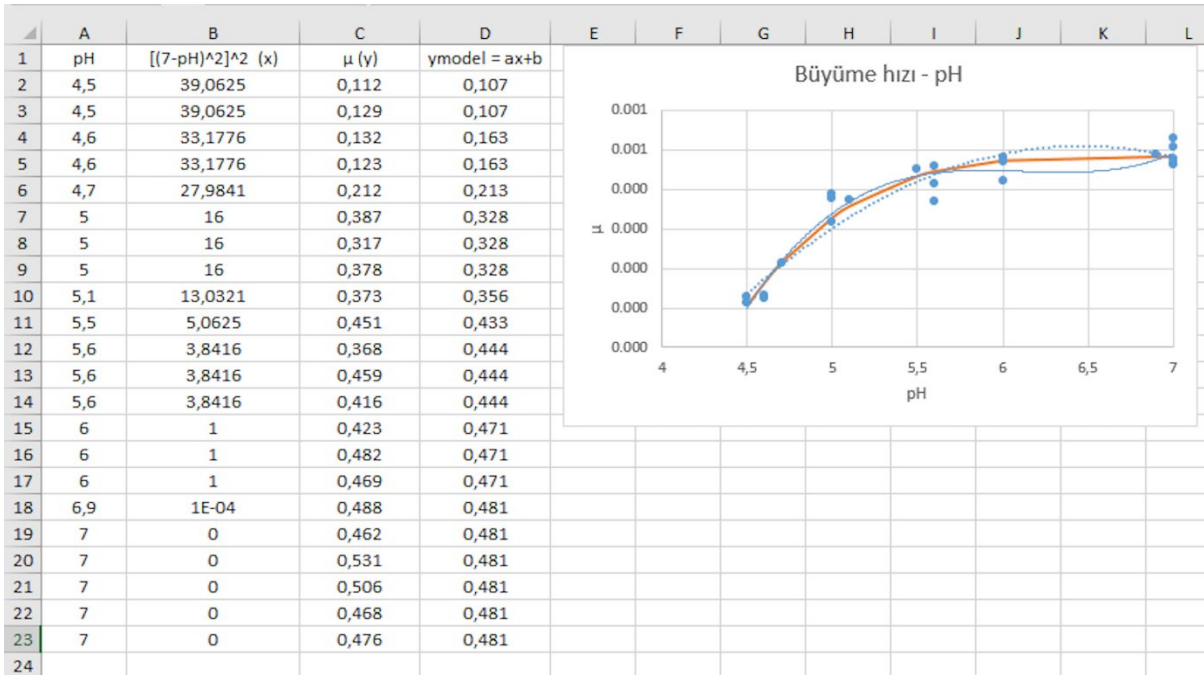
Şekil 10. Bir bakterinin pH'ya karşılık büyüme hızı (μ) verileri. Orijinal veriler Baranyi ve Roberts (1995)'dan alınmıştır.

Figure 10. pH versus growth rate (μ) data of a bacterium. Original data were from Baranyi and Roberts (1995).



Şekil 11. Şekil 10'da gösterilen verinin Excel'de grafikleştirilmiş hali. Mavi daireler deney verilerini, noktalı mavi çizgi veriye uygulanan 2'nci derece polinom modeli ($y = ax^2 + bx + c$), düz mavi çizgi ise uygulanan 3'ncü derece polinom modeli ($y = ax^3 + bx^2 + cx + d$) belirtmektedir.

Figure 11. Sketching the graph of the data in Excel given in Figure 10. Blue circles represent experimental data, dotted blue lines represent 2nd order polynomial model ($y = ax^2 + bx + c$) and solid blue lines represent 3rd order polynomial model ($y = ax^3 + bx^2 + cx + d$).



Şekil 12. Şekil 10’da gösterilen veriler için Excel’de alternatif modelin yazılması ve uyumu. Mavi daireler deney verilerini, noktalı mavi çizgi veriye uygulanan 2’nci derece polinom modeli ($y = ax^2 + bx + c$), düz mavi çizgi uygulanan 3’ncü derece polinom modeli ($y = ax^3 + bx^2 + cx + d$), turuncu kalın çizgi ise alternatif modeli belirtmektedir.

Figure 12. Formulating the alternative model for the data given in Figure 10 in Excel. Blue circles represent experimental data, dotted blue lines represent 2nd order polynomial model ($y = ax^2 + bx + c$), solid blue lines represent 3rd order polynomial model ($y = ax^3 + bx^2 + cx + d$) and thick orange lines represent the alternative model.

Etik Standart ile Uyumluluk

Çıkar çatışması: Yazarlar bu yazı için gerçek, potansiyel veya algılanan çıkar çatışması olmadığını beyan etmişlerdir.

Etik izin: Araştırma niteliği bakımından etik izin gerektirmemektedir.

Finansal destek: -

Teşekkür: -

Kaynaklar

Baranyi, J., Roberts, T.A. (1995). Mathematics of predictive food microbiology. *International Journal of Food Microbiology*, 26, 199-218.

[https://doi.org/10.1016/0168-1605\(94\)00121-L](https://doi.org/10.1016/0168-1605(94)00121-L)

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 in Biomedicine*, 65, 191-200.

[https://doi.org/10.1016/S0169-2607\(00\)00124-3](https://doi.org/10.1016/S0169-2607(00)00124-3)

Davey, K.R., Amos, S.A. (2002). Letter to the editor. *Journal of Applied Microbiology*, 92, 583-584.

<https://doi.org/10.1046/j.1365-2672.2002.1617a.x>

Dolan, K.D., Mishra, D.K. (2013). Parameter estimation in food science. *The Annual Review of Food Science and Technology*, 4, 401-422.

<https://doi.org/10.1146/annurev-food-022811-101247>

Granato, D., Calado, V.M.A., Jarvis, B. (2014). Observations on the use of statistical methods in food science and technology. *Food Research International*, 55, 137-149.

<https://doi.org/10.1002/9781118434635>

Hassani, M., Álvarez, I., Raso, J., Condón, S., Pagán, R. (2005). Comparing predicting models for heat inactivation of *Listeria monocytogenes* and *Pseudomonas aeruginosa* at different pH. *International Journal of Food Microbiology*, 100, 213-222.

<https://doi.org/10.1016/j.ijfoodmicro.2004.10.017>

Jarvis, B. (1989). Statistical aspects of the microbiological analysis of foods. In: Progress in Industrial Microbiology, Vol. 21. Elsevier, Amsterdam. ISBN: 978-0128039748

Montgomery, D.C., Runger, G.C. (2011). Applied statistics and probability for engineers (5th ed.) New York: Wiley. ISBN: 978-0470053041

Moody, H.W. (1982). The evaluation of the parameters in the van Deemter equation. *Journal of Chemical Education*, 59, 290-291.

<https://doi.org/10.1021/ed059p290>

Mossel, D.A.A., Corry, J.E.L., Struijck, C.B., Baird, R.M. (1995). Essentials of the Microbiology of Foods: A Textbook for Advanced Studies. John Wiley & Sons, Chichester. ISBN: 978-0471930365

Ratkowsky, D.A. (2004). Model fitting and uncertainty, in RC McKellar, X Lu (Eds) Modeling Microbial Responses in Food, Boca Raton FL, CRC Press, pp. 151-196. ISBN: 978-0367394653

<https://doi.org/10.1201/9780203503942.ch4>

Ray, B. (2014). Fundamental Food Microbiology. Boca Raton FL, CRC Press, pp. 346. ISBN: 978-1466564435

van Boekel, M.A.J.S. (1996). Statistical aspects of kinetic modeling for food science problems. *Journal of Food Science*, 61, 477-86.

<https://doi.org/10.1111/j.1365-2621.1996.tb13138.x>

van Boekel, M.A.J.S. (2008). Kinetic modeling of food quality: A critical review. *Comprehensive Review in Food Science and Food Safety*, 7, 144-158.

<https://doi.org/10.1201/9781420017410>

van Boekel, M.A.J.S., Zwietering, M.H. (2007). Experimental design, data processing and model fitting in predictive microbiology. In: Modelling Microorganisms in Food, Brul, S., Van Gerwen, S., Zwietering, M.H. (Eds.), pp. 22-43. Woodhead Publishing Ltd: Cambridge, United Kingdom. ISBN: 978-1845690069

<https://doi.org/10.1533/9781845692940.1.22>