



## ARAŞTIRMA

F.Ü.Sağ.Bil.Tıp.Derg.  
2019; 33 (3): 175 - 181  
http://www.fusabil.org

Ahmet Kadir ARSLAN<sup>1, a</sup>  
Zeynep TUNÇ<sup>1, b</sup>  
Cemil ÇOLAK<sup>1, c</sup>

<sup>1</sup> İnönü Üniversitesi,  
Tıp Fakültesi,  
Biyostatistik ve Tıp Bilişimi,  
Malatya, TÜRKİYE

<sup>a</sup> ORCID: 0000-0001-8626-9542  
<sup>b</sup> ORCID: 0000-0001-7956-9772  
<sup>c</sup> ORCID: 0000-0001-5406-098X

### Veri Dönüşümü İçin Açık Kaynak Erişimli Web Tabanlı Yazılım: Veri Dönüşüm Yazılımı<sup>\*</sup>

**Amaç:** Bu çalışmada parametrik test koşulları varsayımlarını sağlamayan veri setlerini çeşitli matematiksel dönüşümler uygulayarak bu varsayımları sağlayabilen veri setlerine dönüştürebilmek için yeni kullanıcı dostu bir web tabanlı yazılım geliştirmek amaçlanmıştır.

**Gereç ve Yöntem:** Bu web tabanlı yazılımı geliştirmek için açık kaynaklı bir R paketi olan Shiny, kullanıldı. Geliştirilen web tabanlı yazılımda  $\arcsinh(x)$ , Box-Cox, Üstel, Lambert W (h tipi), Lambert W (s tipi), Logaritmik, Karekök, Yeo-Johnson veri dönüşüm yöntemleri kullanılmıştır. Ayrıca yazılım dönüşüme tabi tutulacak veri seti için kullanılan yöntemler içinde en iyi yönteme karar vermemizi sağlayacak olan Pearson P test istatistiğini hesaplayarak en iyi dönüşüme karar vermemizi sağlar. Bu işlemi en küçük Pearson P istatistiğine sahip dönüşümü seçerek yapmaktadır.

**Bulgular:** Yazılımın çıktılarını değerlendirebilmek adına üç değişkenden oluşan her bir değişkenin sırasıyla üstel, gamma ve Cauchy dağılımları ile elde edildiği ve değişkenlerin 1000 gözlem içerdiği bir veri seti kullanılmıştır.

**Sonuç:** Geliştirilen yazılım kullanıcıların yapacakları çalışmalarında daha doğru sonuçlar elde etmesini sağlayan ve daha güçlü testler olarak bilinen parametrik test varsayımlarını elde etmeyi sağlayan veri dönüşümü yöntemlerini içeren yeni kullanıcı dostu bir web tabanlı yazılımdır.

**Anahtar Kelimeler:** Normal dağılım, veri dönüşümü, web tabanlı yazılım

### Open Source Access Web Based Software for Data Transformation: Data Transformation Software

**Objective:** In this study, it is aimed to develop a new user-friendly web-based software in order to transform data sets that do not meet the assumptions of parametric test conditions by applying various mathematical transformations into data sets that can provide these assumptions.

**Materials and Methods:** It was used to develop the web based software Shiny, which is an open source R package. In web-based software,  $\arcsinh(x)$ , Box-Cox, Exponential, Lambert W (h type), Lambert W (s type), Logarithmic, Square root, Yeo-Johnson data transformation methods were used. The software also allows us to decide on the best conversion for the data set to be transformed by calculating the Pearson P test statistic, which will allow us to decide on the best method in the methods used. This is done by selecting the conversion with the smallest Pearson P statistic

**Results:** In order to evaluate the printouts of the software, a data set consisting of three variables obtained by the contribution of Üstel, Gamma and Cauchy including 1000 observations was used.

**Conclusion:** The developed software is a new user-friendly web-based software that enables users to get more accurate results in their work and to obtain data transformation methods that provide parametric test assumptions known as more powerful tests.

**Key words:** Normal distribution, data transformation, web-based software

Geliş Tarihi : 22.05.2019  
Kabul Tarihi : 23.10.2019

### Giriş

Araştırmacılar tarafından toplanan verilerle istatistiksel çıkarımlar yaparak amaca uygun sonuçlara varılabilmesi için istatistiksel hipotez testlerinden yararlanılır. Veri analiz etmede kullanılacak istatistiksel hipotez testleri; verinin ölçek türüne ve yapısına, dağılımın biçimine ve test edilecek hipoteze göre "parametrik testler" ve "parametrik olmayan testler" olarak sınıflandırılmaktadır. Parametrik testler belirli bir hipotetik evren dağılımına ve bu dağılımı belirleyen ortalama, standart sapma gibi bazı parametrelere dayalı olup esnek olmayan yöntemlerdir. Parametrik olmayan testler ise genellikle bir evren dağılımından bağımsız olarak, beklenen-gözlenen değer farkları, sıra ve sıra farkları gibi daha temel ve betimsel kestirime dayalı esnek yöntemlerdir (1). Parametrik testlerin istatistiksel gücü ve etki büyüklüğü, parametrik olmayan testlere göre daha yüksektir. Ek olarak birinci ve ikinci tip hatalara karşı parametrik testler daha dirençlidir (2). Parametrik testler belirli bir dağılım ve parametreler esas alınarak

### Yazışma Adresi Correspondence

Zeynep TUNÇ

İnönü Üniversitesi,  
Tıp Fakültesi,  
Biyostatistik ve Tıp Bilişimi  
Malatya - TÜRKİYE

zeynep.tunc@inonu.edu.tr

<sup>\*</sup> 3. Uluslararası Multidisipliner Araştırmalar ve Yenilikçi Teknolojiler Sempozyumu, 11-13 Ekim 2019, Ankara / Türkiye.

geliştirildiği için bu sınırlılıklara bağlı olarak tanımlanan bazı varsayımlar içermekte ve sadece bu varsayımlar sağlandığında kullanılabilir. Bu varsayımlar örneklemelerin çekildiği evrenlerin normal dağılım göstermesi ve evren varyanslarının homojen olmasıdır (3). Varsayımların ihmal edilmesi test sonuçlarının hatalı ve yanıltıcı olmasına neden olmaktadır. Bu nedenle veri analizi yapmadan önce kullanılacak teste karar vermek için varsayımlar kontrol edilmeli ve parametrik ya da parametrik olmayan testlerden hangisinin seçileceğine karar verilmelidir (4). Bu nedenle öncelikle varsayımlardan biri olan verilerin normal dağılıma uygunluğu araştırılmalıdır. Eğer veriler normal dağılıma uyuyor ve diğer varsayımlar sağlanıyor ise istatistiksel analizlerde parametrik testler kullanılmalıdır. Uygulanan analizler sonucunda verilerin normal dağılıma uygunluğu sağlanmıyorsa yada varyansların homojenliği koşulu sağlanmıyorsa normal dağılımı ve homojen varyans yapısını bozduğu düşünülen verilere bazı dönüşümler yapılarak veri setinin normal dağılım göstermesi ve varyansların homojen dağılımı sağlanabilir. Böylece yeni elde edilen veri setine parametrik testler uygulanabilir. Bu sayede parametrik olmayan testler yerine istatistiksel gücü ve etki büyüklüğü daha kuvvetli olan birinci ve ikinci tip hatalara karşı daha dirençli olan parametrik testler ile veri setleri analiz edilerek çalışmanın gücü artırılmış olur (2).

Araştırmacılar arasında genellikle test seçiminde parametrik testleri seçme ve bu testleri uygulama eğilimi vardır (5). Çalışmalarda bazen parametrik test seçimine karar verilmesini sağlayan en önemli varsayımlardan biri olan normallik veya varyansların homojenlik varsayımı göz ardı edilmekte ve yanlış test seçimi yapılmaktadır. Bazen de matematiksel dönüşümler göz ardı edilerek aslında uygulanabilecek matematiksel dönüşümlerle normal dağılımı sağlayabilecek veri setlerine yanlış testler uygulanmaktadır. Bu gibi durumlara karşılaşılmaması için öncelikle sayılan varsayımlar kontrol edilmeli ve bu varsayımlar sağlanmıyor ise veri setinde yer alan değişkenlere bazı matematiksel dönüşümler uygulanarak yeni elde edilen veri setinin normal dağılıma uygunluğu kontrol edilmelidir. Bu kontroller ve işlemler gerçekleştirildikten sonra istatistiksel test seçimi yapılmalıdır. Bu çalışma kapsamında istatistiksel analizlere başlamadan önce daha test seçimi aşamasındayken; hangi hipotez testini seçeceğimize karar vermemizi sağlayacak en önemli koşullardan olan normal dağılıma uygunluk ve varyansların homojenliği koşulunu kontrol ettikten sonra sağlamayan veri setlerini, normal dağılım koşulunu ve homojen varyanslılık koşulunu sağlayan veri setlerine dönüştürebilmek amacıyla kullanılacak veri dönüşüm tekniklerini içeren yeni kullanıcı dostu bir web tabanlı yazılım geliştirmek amaçlanmıştır.

## Gereç ve Yöntem

**Veri seti:** Bu çalışmada yapılan yazılımın çalışma şeklini gösterebilmek ve çıktılarını değerlendirebilmek adına üç değişkenden oluşan her bir değişkenin

sırasıyla üstel, gamma ve Cauchy dağılımları ile elde edildiği ve değişkenlerin 1000 gözlem içerdiği veri seti R Project versiyon (3.4.2) yazılımı (6) ile türetilmiştir.

**Veri Dönüşüm Yöntemleri:** Yapılacak çalışmalarda araştırmacılar genelde istatistiksel gücü ve etki büyüklüğü daha kuvvetli olan birinci ve ikinci tip hatalara karşı daha dirençli olan parametrik testleri seçme eğiliminde olup bu testleri seçmek adına bazı varsayımların gerçekleşip gerçekleşmediğini kontrol etmelidir. Bu varsayımlar örneklemelerin çekildiği evrenlerin normal dağılım göstermesi ve evren varyanslarının homojen olması varsayımlarıdır (3). Araştırmacılar bu nedenle öncelikle varsayımlardan biri olan verilerin normal dağılıma uygun olup olmadığını araştırmalıdır. Eğer veriler normal dağılıma uyuyor ve diğer varsayımlar sağlanıyor ise istatistiksel analizlerde parametrik testler kullanılmalıdır. Fakat veriler normal dağılıma uygunluk varsayımını sağlamıyor ya da varyansların homojenlik koşulunu sağlamıyor ise normal dağılıma uymadığı belirlenen ve homojenliği bozduğu düşünülen değişkenlere bazı dönüşümler yapılarak veri setinin normal dağılıma uygunluğu sağlanabilir. Yapılan bu dönüşümlere veri dönüşüm yöntemleri denir. Geliştirilen web tabanlı yazılımda  $\text{arcsinh}(x)$ , Box-Cox, Üstel, Lambert W (h tipi), Lambert W (s tipi), Logaritmik, Karekök, Yeo-Johnson veri dönüşüm yöntemleri kullanılmıştır. Ayrıca yazılım dönüşüme tabi tutulacak veri seti için kullanılan veri dönüşümü yöntemleri içinde en iyi yöntem karar vermemizi sağlayacak olan Pearson P test istatistiğini hesaplayarak en iyi dönüşüm ile dönüştürülen veri setini indirmemize olanak sağlamaktadır. Bu işlemi yazılımda yer alan veri dönüşümü yöntemlerinin her biri için elde edilen Pearson P istatistiği değerlerinden en küçüğüne sahip dönüşümü seçerek yapmaktadır.

**Logaritmik dönüşüm:** Sağa ya da sola çarpık olup değişim aralığının ve merkezi simetrisinin önemli olduğu veri setlerinde varyansa dengeleme dönüşümü olarak kullanılan bir veri dönüşüm yöntemidir. Dönüşümlerde en yaygın olarak 10 tabanındaki logaritmik dönüşüm tercih edilirken, doğal logaritma olarak adlandırılan e tabanlı logaritmik dönüşüm de kullanılmaktadır. Eğer veri seti sıfır ve negatif değerler içeriyorsa veri setinde yer alan tüm gözlemlere pozitif bir sayı eklenerek veri seti düzenlenir ve daha sonra dönüşüm yapılır. Kesikli bir veri seti ile ilgileniliyor ve veri setinde sıfır değerleri varsa her bir gözleme 0.5 değeri eklenerek dönüşüm yapılmalıdır. Veri setindeki tüm değerlerin logaritması alınarak işlemler yapılır ancak sonuçlar yorumlanırken veriler eski değerlerine dönüştürülmelidir (7).

**Üstel dönüşüm:** Veri setinde yer alan tüm gözlemlerin p. inci kuvvetinin alınarak yapıldığı dönüşümdür. Bu dönüşüm uygulanırken verinin kaçınıcı kuvvetinin alınacağına şu şekilde karar verilir:

Veri setindeki her bir değişken için  $3.kartil - 1.kartil = IQR$  değerleri hesaplanarak logaritması alınır ve karşılık gelen medyan (M) değerlerinin de logaritması alınarak saçılım grafiği çizilir. Buradan  $\log(IQR) = a + b.log(M)$  denkleminde ait olan eğim

katsayısından faydalanılarak  $b = 1 - p$  olacak şekilde  $p$  değeri belirlenir (8).

**Karekök dönüşümü:** Üstel dönüşüm ile elde edilen  $p$  değerinin  $\frac{1}{2}$  ye eşit olması durumunda üstel dönüşüm özel hal olarak karekök dönüşümü adını alır. Bu dönüşüm normalleştirme durumunda kullanılabilmesi gibi varyans dengeleme durumunda da kullanılmaktadır (9). Karekök alma işleminden dolayı içinde negatif gözlem olan veri setlerinde bu dönüşümü kullanmadan önce her bir gözleme sabit bir sayı eklenerek veri setindeki tüm gözlemlerin sıfır ve pozitif olması sağlanmalıdır (8).

**Arcsinh(x) dönüşümü:** Matematikte hiperbolik fonksiyonlar bilinen trigonometrik fonksiyonların analogudur. Temel hiperbolik fonksiyonlar **hiperbolik sinüs** "sinh", **hiperbolik kosinüs** "cosh" olarak bilinir. Diğer fonksiyonlar bu iki fonksiyondan türetilmektedir. Hiperbolik fonksiyonlar genellikle difereansiyel denklemlerin çözümünde önemli rol oynayan fonksiyonlardır.

$$\sinh(x) = \frac{e^x - e^{-x}}{2}$$

ve

$$\cosh(x) = \frac{e^x + e^{-x}}{2}$$

dir. Hiperbolik fonksiyonların bire bir oldukları aralıklarda ters fonksiyonları bulunur. Bu yazılımda sinh fonksiyonunun tersi olan arcsinh fonksiyonu veri dönüşümü yöntemi olarak kullanılmıştır.

$$\operatorname{arcsinh}(x) = \ln(x + \sqrt{x^2 + 1})$$

ile dönüşümler yapılmıştır (10).

**Box-Cox dönüşümü:** Box ve Cox (11) tarafından 1964 yılında yazılan makalelerinde yer verilen Box-Cox dönüşümü yalnızca tek bir parametre olan  $\lambda$  ya bağlı olup basit bir dönüşümdür.  $\lambda$  ya bağlı olarak Box-Cox dönüşümü

$$g(x; \lambda) = 1_{\lambda \neq 0} \frac{x^\lambda - 1}{\lambda} + 1_{\lambda=0} \log x$$

Burada  $x$  orijinal yani işlem yapılmamış veri setindeki gözlemleri temsil etmektedir.  $\lambda$  parametresi maksimum olabilirlik ile tahmin edilebilir.

**Yeo-Johnson dönüşümü:** 2000 yılında Yeo ve Johnson (12) tarafından önerilen Yeo-Johnson dönüşümü, normal dağılım ile dönüştürülmüş dağılım arasındaki Kullback-Leibler mesafesini en aza indiren lambda değerini (aşağıdaki denklemde) bulmaya çalışır.

$$g(x; \lambda) = 1_{(\lambda \neq 0, x \geq 0)} \frac{(x+1)^\lambda - 1}{\lambda} + 1_{(\lambda=0, x \geq 0)} \log(x+1) \\ + 1_{(\lambda \neq 2, x < 0)} \frac{(1-x)^{2-\lambda} - 1}{\lambda - 2} \\ + 1_{(\lambda=2, x < 0)} \log(1-x)$$

Box-Cox da yer alan  $\lambda$  parametresi gibi bu formülde yer alan  $\lambda$  parametresi de maksimum olabilirlik ile tahmin edilebilir.

**Lambert W (h tipi) ve Lambert W (s tipi) dönüşümü:** Georg ve Lambert (13) tarafından Lambert W paketinde yer alan Lambert WxF dönüşümü momentleri kullanarak rastgele bir  $X$  değişkenini çözen bir mekanizmadır. Bu yöntem, Box-Cox veya Yeo-Johnson dönüşümü kadar iyi performans göstermez. Bu dönüşümler çarpık (skewed), çok çarpık (heavy-tailed) ve çok ağır çarpık (skewed heavy-tailed) olarak üç tür dönüşümü gerçekleştirir. Geliştirilen yazılımda bu dönüşümlerden çarpık (skewed) ve çok çarpık (heavy-tailed) durumuna yer verilmiştir.

**Geliştirilen web tabanlı yazılım:** Web tabanlı bu uygulamayı oluşturmak için, R programlama dili temelinde interaktif web tabanlı uygulamaların tasarlanmasına izin veren Shiny kütüphanesi kullanılmıştır (14). Geliştirilen yazılım ayrıca İngilizce dil seçeneğini de içermektedir. Yazılıma ait ana ve alt menüler aşağıda açıklanmıştır.

**Veri İşlemleri:** Bu web tabanlı uygulamanın ilk aşamasında, veri kümesini içeren dosya yüklenir. Veri analizinde farklı uzantılara sahip en yaygın kullanılan dosya türlerinden olan MS Excel (.xls / .xlsx) ve SPSS (.sav) dosya türleri ile yükleme yapılır.

**Dönüşüm:** Web tabanlı uygulamanın bu menüsünde 'Sayısal değişken(ler)' sekmesi ile veri setinde yer alan parametrik test yapabilmek adına kontrol edilmesi gereken varsayımlardan olan normallik ve homojen varyanslılık koşullarını bozduğu belirlenen değişkenler seçilir. Seçilen değişkenler için bu sekmede yer alan çalıştır seçeneği kullanılır. Bu şekilde yazılımda yer verilen veri dönüşümü yöntemleri kullanılarak seçilen değişkenler için yazılım gerekli analizleri yapar. Seçilen değişkenler için dönüşümler yapılırken aynı zamanda uygulanan tüm yöntemler için en iyi yönteme karar vermemizi sağlayacak olan Pearson P test istatistiği hesaplanır ve en küçük Pearson P istatistiğine sahip yöntem seçilir. Ayrıca seçilen yöntem ile dönüşüme tabi tutulan yeni veri seti ve dönüşümden önce var olan veri seti için yoğunluk grafikleri de verilmektedir. Bu çıktılar bu menüde yer verilen "Sayfayı yazdır" butonu ile yazdırılabilir. Şekil 1 "Dönüşüm" menüsünü göstermektedir.

Sayısal değişkenler:

sepal length, sepal width, petal length, petal width

Yeniden örnekleme yöntemi kullanılmı mı?

**Çalıştır**

Şekil 1. "Dönüşüm" menüsü

Analiz yapmak istediğiniz veri dosyasını yükleyiniz.

Seçiniz... veri\_trans.xlsx

**Yükle**

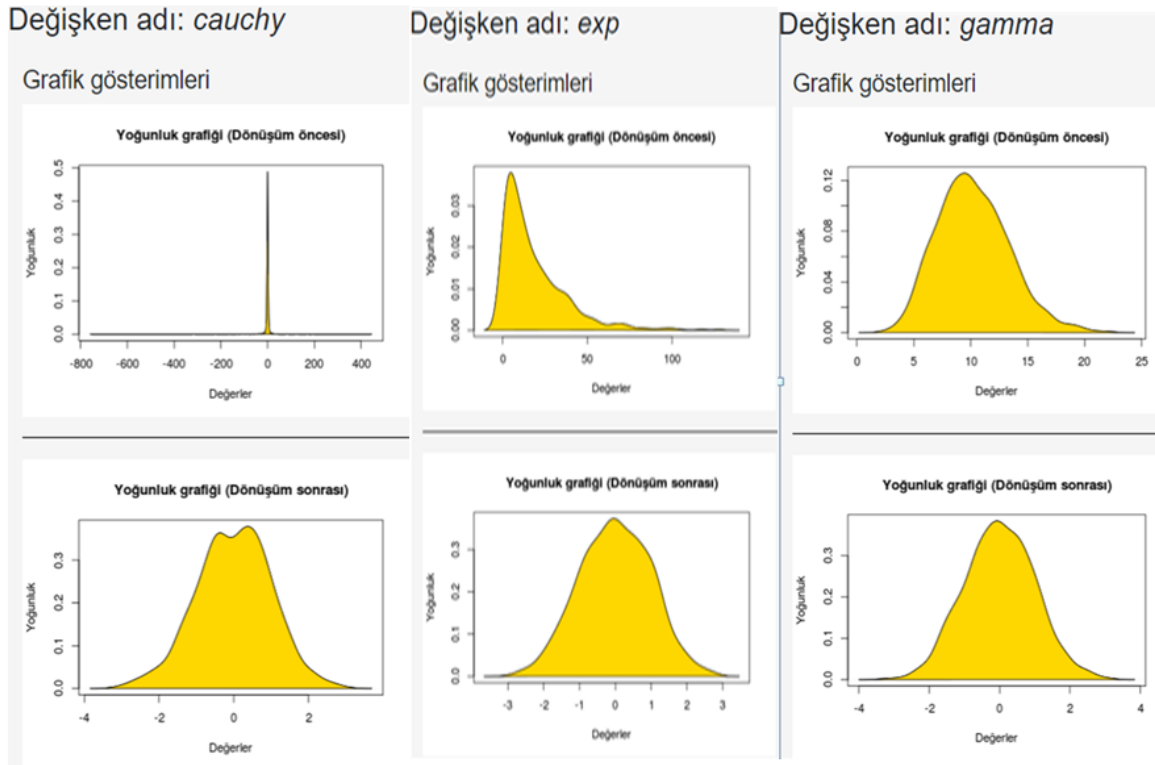
Sayfada 10 kayıt göster

Değişken	Tipi	Rolü
cauchy	Sürekli Sayısa	Tahminleyici
exp	Sürekli Sayısa	Tahminleyici
gamma	Sürekli Sayısa	Tahminleyici

Önceki 1 Sonraki

cauchy	exp	gamma
-1.64168422737822	16.1281836981659	12.7815499141471
-0.128484752175862	44.1101594616599	10.3936557788716
-5.14419178028534	14.2787746890428	10.9464943033326
-0.349149100169962	21.5979083805325	12.1925618577977
-0.222143526199769	32.9128254223936	5.1279443522741
-0.528816537126802	6.61970895715058	9.543260415571
1.8693018582998	20.1357735871317	11.5890413720304
24.1257842341142	12.5500777270645	6.85263150600538
-6.21392811621518	25.6188890154691	13.9930294077382
-27.7018732249839	0.83195210847881	12.4530480811778
0.974474914528647	28.6898336148347	11.1039067339745
-222.353349107748	0.0315885460009083	7.0516578525199
-0.18409418359269	46.5132224132255	6.82263690562301
6.53358868767386	1.41852992586792	9.48592570250217
-0.0100052737411257	18.821991457695	12.4430341404023

Şekil 2. Yüklenen dosyaya ilişkin görüntüler



Şekil 3. Dönüşüm öncesi ve sonrası yoğunluk grafiklerine ilişkin görüntüler

Değişken adı: <i>cauchy</i>			Değişken adı: <i>exp</i>			Değişken adı: <i>gamma</i>		
Dönüşüm türü	ƒ	İstatistik	Dönüşüm türü	ƒ	İstatistik	Dönüşüm türü	ƒ	İstatistik
arcsinh(x)		3.05	arcsinh(x)		4.021	arcsinh(x)		1.269
Üstel		34.483	Box-Cox		0.726	Box-Cox		0.492
Lambert W (h-tipi)		1.053	Üstel		1064.626	Üstel		983.004
Lambert W (s-tipi)		342.9	Lambert W (h-tipi)		23.362	Lambert W (h-tipi)		1.852
Logaritmik		922.887	Lambert W (s-tipi)		1.126	Lambert W (s-tipi)		0.523
Dönüşüm yok		393.866	Logaritmik		4.643	Logaritmik		1.331
Karekök		539.09	Dönüşüm yok		23.932	Dönüşüm yok		1.713
Yeo-Johnson		281.891	Karekök		3.438	Karekök		0.437
			Yeo-Johnson		1.24	Yeo-Johnson		0.497
<b>Seçilen en uygun dönüşüm türü:</b> <i>Lambert W (h-tipi)</i>			<b>Seçilen en uygun dönüşüm türü:</b> <i>Box-Cox</i> (lamda=0.2651)			<b>Seçilen en uygun dönüşüm türü:</b> <i>Karekök</i>		

Şekil 4. Dönüşüm yöntemlerine ait Pearson p istatistiği tablosu

Ek olarak bu menüde yazılım çalıştırıldıktan sonra seçilen en iyi yöntemle dönüştürülmüş veri setinin MS Excel (.xls / .xlsx) formatında indirilmesini sağlayan "Dönüştürülmüş veriyi indir" butonu vardır. Veri dönüşümü sonrası daha tutarlı Pearson P istatistiği hesaplayabilmek için "tekrarlı 5-katlı çapraz geçerlilik" ve "birini dışında bırakma çapraz geçerlilik" yeniden örnekleme yöntemleri yazılımda kullanıcılara seçmeli olarak sunulmaktadır. Tekrarlı n-katlı çapraz geçerlilik yönteminde veriler öncelikle n parçaya ayrılır ve kullanılan model n parça için uygulanır. n parçadan bir tanesi test için kullanılırken, diğer n-1 parça modelin eğitimi için kullanılır. Elde edilen değerlerin ortalaması, çapraz geçerlilik yöntemi için değerlendirilir (15). Birini dışında bırakma çapraz geçerlilik yöntemi ise eğitim veri setinde ayırdığı bir gözlem hariç diğer gözlemleri kullanıp ayırdığı gözlem üzerinde test verisini değerlendirmektedir (16).

**Geliştirilmiş interaktif web uygulamasının erişilebilirliği ve alıntılanması:** Geliştirilen web tabanlı yazılıma <http://biostatapps.inonu.edu.tr/VDY/> adresinden ücretsiz olarak erişilebilir. Bilimsel çalışmalarda yazılımın kaynak gösterilmesi "Arslan, A. K. , Tunc, Z. & Colak, C. VDY: Veri Dönüşüm Yazılımı [Web-tabanlı yazılım]" ile yapılmaktadır. Bu yazılımın grafik ara yüzünün geliştirilmesinde shiny (14) ve shinydashboard (17) kütüphaneleri, arka planda çalışan uygulamalarda ise bestNormalize (18) kütüphanesi kullanıldı.

## Bulgular

Yazılımın çalışma şeklini gösterebilmek ve çıktılarını inceleyebilmek için üç değişkenden oluşan her bir değişkenin sırasıyla üstel, gamma ve Cauchy dağılımları ile elde edildiği ve değişkenlerin 1000 gözlem içerdiği veri seti üzerinde veri dönüşümü yöntemleri uygulanmıştır. Veri setine ilişkin tanımlayıcı istatistik tablosu tablo 1 de verilmiştir.

Tablo 1. Değişkenlere ait tanımlayıcı istatistik tablosu

Değişkenler	Değişken Türü	Ortalama±Standart sapma
X <sub>1</sub>	Nümerik	-0.0004±0.98773
X <sub>2</sub>	Nümerik	49.8250±7.01824
X <sub>3</sub>	Nümerik	-0.0205±1.01109

İlk olarak veri seti yazılıma yüklenmiştir. Şekil 2 yüklenen dosyaya ilişkin görüntüleri göstermektedir.

Sonra "Dönüşüm" sekmesi ile veri dönüşümü yapılmak istenen değişkenler "sayısal değişkenler" sekmesi ile seçilir. İstenilirse yeniden örnekleme yöntemleri de kullanılarak "çalıştır" butonu ile yazılım çalıştırılır. Şekil 3 veri dönüşümü işlemi sonuçlarına ait çıktılarından değişkenlere ait dönüşüm öncesi ve sonrası yoğunluk grafiklerini göstermektedir.

Bu grafiklere göre normal dağılıma uygun olmayan grafiklere sahip değişkenlerin dönüşüm işleminden sonra normal dağılıma uygun olacak şekilde yeni grafiklere sahip olduğu gözlenmektedir.

Şekil 4 ise yazılımda yer verilen değişken dönüşüm yöntemlerinin her bir değişkene uygulanmasıyla elde edilen Pearson p istatistiğine ait tabloyu ve her bir değişken için en uygun yöntemin hangisi olduğunu gösteren nota ait görüntüleri göstermektedir.

Bu tabloya göre hesaplanan Pearson p istatistiği değerlerine göre değişkenler için en küçük Pearson p istatistiğinin elde edildiği dönüşüm o değişken için en iyi dönüşüm olup, yöntemler sırasıyla Lambert W (h tipi), Box-Cox ve karekök dönüşümleridir.

## Tartışma

Yapılacak çalışmalarda veri setlerine uygulanacak istatistiksel yöntemlere karar verilirken yani parametrik ya da parametrik olmayan hipotez testlerinden hangisinin seçileceğini belirlerken verilerin normal dağılımı sağlaması ve varyansların homojen olması varsayımları kontrol edilmelidir. Eğer çalışmanın yapılacağı veri seti parametrik hipotez testi koşulları olan bu varsayımları sağlamıyorsa genellikle parametrik olmayan hipotez testleri ile analizler yapılır (19). Fakat parametrik olmayan hipotez testleri istatistiksel gücü ve etki büyüklüğü daha kuvvetli olan birinci ve ikinci tip hatalara karşı daha dirençli olan parametrik hipotez testlerine göre daha zayıf sonuçlar vermektedir (2). Bu nedenle araştırmacılar test seçimi aşamasında eğer parametrik test koşulları sağlanmadığı sonucunu elde ediyorsa ilk tercihleri parametrik olmayan testleri kullanmak yerine veri dönüşüm yöntemlerinden faydalanmak olmalıdır. Bu şekilde veri seti için kullanılan veri dönüşümü yöntemleri ile veri setinin varsayımları sağlayabilecek bir konuma getirilmesi sağlanabilir. Böylece yeni veri seti ile daha güçlü istatistiksel bilgiler elde edebileceğimiz parametrik testleri kullanmış oluruz (20).

Bu nedenle araştırmacıların veri dönüşümü işlemlerini kolayca yapabileceği ve farklı yöntemler arasından seçim yapabildiğini sağlayan bir kullanıcı dostu web yazılımı geliştirilmiştir. Geliştirilen yazılımda sekiz adet veri dönüştürme yöntemi kullanılmıştır. Yazılımda yer verilen her bir yöntem için Pearson P istatistikleri hesaplanarak dönüşüme tabi tutulacak veri seti için iyi yöntem karar vermemiz sağlanmıştır. Bu

işlemi yazılımda yer alan veri dönüşümü yöntemlerinin her biri için elde edilen Pearson P istatistiği değerlerinden en küçüğüne sahip dönüşümü seçerek yapmaktadır.

Bilinen paket programlardan olan IBM SPSS Statistics (21), Minitab (22), MedCalc (23)'da veri dönüşümü işlemlerinin yapılmasına imkan sağlayan bir modül yoktur. Fakat geliştirilmiş web tabanlı yazılımda veri dönüştürme işlemleri kolayca yapılabilir. Araştırmacılar ayrıca geliştirilen bu yazılım ile yazılımda yer verilen dönüşüm yöntemlerinden en iyisi ile dönüştürülerek elde edilen yeni veri setini MS Excel (.xls / .xlsx) formatında indirebilirler ve bu sayede yeni veriyi direk çalışmalarında kullanabilirler. Geliştirilen yazılım ek olarak en iyi yöntemi belirlemek için bulunan Pearson P istatistiği değerlerini daha tutarlı hesaplayabilmek için "tekrarlı 5-katlı çapraz geçerlilik" ve "birini dışında bırakma çapraz geçerlilik" yeniden örnekleme yöntemlerini kullanıcılara sunmaktadır.

Ek olarak bu çalışmada geliştirilen yazılımın çıktılarını değerlendirebilmek adına üç değişkenden oluşan her bir değişkenin sırasıyla üstel, gamma ve Cauchy dağılımları ile elde edildiği ve değişkenlerin 1000 gözlem içerdiği veri seti R Project versiyon (3.4.2) yazılımı ile türetilmiştir. Analiz sonuçlarına göre 3 değişken içinde normal dağılıma uygunluğun sağlanabilmesi adına yapılan veri dönüşümü yöntemlerine göre elde edilen Pearson p istatistikleri hesaplanmıştır. Hesaplanan istatistiğin en küçük olduğu dönüşüm yöntemi o değişken için en uygun yöntemi belirlemektedir. Sırasıyla 3 değişkenli veri seti için en uygun yöntemler Lambert W (h tipi), Box-Cox ve karekök dönüşümleri şeklinde elde edilmiştir.

## Kaynaklar

1. Buskirk TD, Willoughby LM, Tomazic TJ. Nonparametric statistical techniques. The Oxford Handbook of Quantitative Methods 2013; 2: 106-141.
2. Kartal M. Bilimsel Araştırmalarda Hipotez Testleri. Ankara: Nobel Yayın Dağıtım, 2010.
3. Alpar R. Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş. Ankara: Detay Yayıncılık, 2013.
4. Field A. Discovering statistics using IBM SPSS statistics. sage, 2013.
5. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965; 52: 591-611.
6. Team RC. R: A language and environment for statistical computing. 2019.
7. McDonald JH. Handbook of biological statistics. Baltimore: MD: Sparky house publishing, 2009.
8. Ovla HD, Taşdelen B. Aykırı değer yönetimi. Mersin Üniversitesi Sağlık Bilimleri Dergisi 2012; 5: 1-8.
9. Petrie A, Sabin C. Medical statistics at a glance. John Wiley & Sons, 2013.
10. Bradley RE, D'Antonio LA, Sandifer CE. Euler at 300: an appreciation: StatSoft, Inc., 2007.
11. Box GE, Cox DR. An analysis of transformations. J Royal Stat Soc Series B Stat Methodol 1964; 26: 211-243.
12. Yeo IK, Johnson RA. A new family of power transformations to improve normality or symmetry. Biometrika 2000; 87: 954-959.
13. Goerg G. LambertW. Probabilistic Models to Analyze and Gaussianize Heavy-Tailed, Skewed Data. R package version 06. 2016;4.
14. Cheng J, Xie Y, McPherson J. shiny: web application framework for R. R package version 0.13. 2. 2016.
15. Mitchell TM. Artificial neural networks. Machine Learning 1997; 45: 81-127.
16. Yeniay O, Goktas A. A comparison of partial least squares regression with other prediction methods. Hacettepe Journal of Mathematics and Statistics 2002; 31: 99-101.
17. Chang W, Borges Ribeiro B. Shinydashboard: Create dashboards with Shiny'. R package version 06. 2017;1.
18. Peterson RA. bestNormalize: A suite of normalizing transformations. R package version 341. 2017.
19. Pasin Ö, Ankaralı H, Cangür Ş, Sungur MA. Parametrik olmayan çok değişkenli varyans analizi ve sağlık alanında bir uygulaması. Bilişim Teknolojileri Dergisi 2016; 9: 13.

20. Karagöz Y. Nonparametrik tekniklerin güç ve etkinlikleri. Elektronik Sosyal Bilimler Dergisi 2010; 9: 18-40.
21. Corp I. IBM SPSS statistics for windows, version 25.0. Armonk, NY: IBM Corp. 2017.
22. Minitab I. MINITAB statistical software. Minitab Release. 2000;13.
23. Schoonjans F, Zalata A, Depuydt C, et al. MedCalc: A new computer program for medical statistics. Comput Methods Programs in Biomed 1995; 48: 257-262.