



İstatistiksel Veri Analizinde R Yazılımı ve Kullanımı* R Software and Its Use in Statistical Data Analysis*

C. Deha Doğan, Ankara Üniversitesi Eğitim Bilimleri Fakültesi, ddogan@ankara.edu.tr
Müge Uluman, Marmara Üniversitesi Eğitim Bilimleri Fakültesi, mugeulumann@gmail.com

ÖZ. Bu çalışmanın amacı son yıllarda kullanımı yaygınlaşan ve diğer istatistik paket programlara kıyasla belirli avantajlara sahip olan R yazılımını temel düzeyde tanıtmak ve Türkiye’de 2010-2014 yılları içerisinde eğitim bilimleri alanında yapılan ve SSCI kapsamında taranan 3 akademik dergide yayınlanan araştırmalardaki R yazılımının ve diğer paket programların kullanım sıklığını belirlemektir. Yaygın olarak kullanılan istatistiksel paket programlarının aksine parasız temin edilebilen ve açık kaynaklı olan R yazılımı dünyanın farklı yerlerindeki araştırmacıların katkıları ile sürekli gelişen aktif bir yapıya sahiptir. Bu yapısı ile hem araştırmacıların bilimsel araştırma sürecinde hem de öğretmenlerin okul uygulamalarında kullanabilecekleri tüm analizlerin hesaplanmasına olanak tanımaktadır. Betimsel tarama türünde olan araştırmada elde edilen veriler incelendiğinde 1627 makale arasında R yazılımının kullanıldığı sadece bir çalışma olduğu, bunun yanı sıra en sık kullanılan istatistiksel paket programların ise SPSS ve Lisrel olduğu belirlenmiştir. Bu bulgular önemli avantajlara sahip olan R yazılımının hemen hemen hiç kullanılmadığı şeklinde yorumlanabilir. R yazılımının kullanımının yaygınlaştırılması için üniversitelerde hizmet içi eğitimlerin gerçekleştirilmesi yerinde olacaktır.

Anahtar Kelimeler: R yazılımı, R CRAN, R dili, İstatistiksel Veri Analiz Yazılımları

ABSTRACT. The aim of this study is two fold. It aims at introducing the R software at basic level as an analysis instrument which has been widespread in recent years and offers certain advantages in comparison to other statistical packages and identifying the frequency of the use of R and other packages in research in educational sciences published within the last 5 years in Turkey in 3 academic journals covered by the SSCI. Unlike the widely used statistical packages, R is an open-source software available free of charge with a dynamic structure constantly developed with the contributions of researchers from all over the world. Such a structure allows calculation of all kinds of statistical analysis which can be used by not only researchers in scientific research process but also teachers in school practices. The study was designed in descriptive survey model. It was found out that the R software was used in only one out of 1627 articles, while SPSS and Lisrel were the most frequently used statistical packages. It can be inferred from these findings that the R software has almost no use despite the significant advantages it offers. Hence, it is recommended to realize in-service training is held at universities in order to promote the use of the R software.

Keywords: R software, R CRAN, R language, Statistical Data Analyses Software

SUMMARY

Purpose and Significance: There are many statistical package software that can be used for analyzing data in quantitative studies. Among those package software R has become popular since it has some advantages. R is a free software environment for statistical computing and graphics. Unlike many other software companies that protectively hide away the code on which their software is based, R is an open source software. In R it is allowed to access the code of a specific analysis. This allows everyone to contribute to the software. So capability of R dynamically expands because people from all over the world contribute to it (Field, Miles, Files, 2012). It is also possible to download some packages for specific statistical analysis. The basic of R software depends on the “S” language developed by Becker and Chambers. It is expanded version of commercial software called S-PLUS (Er & Sönmez).

Beaujean (2013) emphasizes three important advantages of R. One reason is that it is a very powerful programming language that allows conduct a wide range of quantitative analysis. Second reason is that it allows users to submit their own packages to the R server for anyone to use. As a result of this today there has been more than 4000 statistical packages. It is very easy for a user to download the required package and conduct the analysis. A third reason for using R is that it is a free open source.

*Bu çalışmanın bir bölümü Uluslararası Gelecek için Eğitim (ICEFIC 2015) kongresinde sözlü bildiri olarak sunulmuştur.

Although R software has very advantages, it is not commonly used in Turkey for quantitative data analysis unlike some other countries. It is not known well by the researchers in Turkey. The aim of this study is to explain the usage of R in social sciences and define how frequently it is used in scientific studies in Turkey. Below are the research questions;

1. How frequently R software has been used in the quantitative studies published in Turkey?
2. How frequently other commercial statistical software has been used in the quantitative studies published in Turkey?

Methodology: This is a survey type of study which aims to define frequency of statistical software used in quantitative studies published in Turkey. The required data collected from 1627 articles which were published in Educational Science: Theory and Practice, Education and Science, Hacettepe University Educational Faculty Journal from 2010 to 2014. 1627 articles have been examined and categorized according to the statistical software that is used for quantitative data analysis. Then the frequencies and percentages have been calculated for each statistical software. Percentages of using statistical software have been compared according to the years.

Results: Analysis of the data shows that out of 1627 articles published in three journals between 2010-2014 in Turkey, 495 of them are not required to do quantitative data analysis. They are mostly qualitative studies. So the rest 1132 quantitative studies require usage of statistical software. So 1132 articles have been categorized and analyzed according to the statistical software that is used for quantitative data analysis. The outstanding results of the study is that in % 50 of the articles, it is not presented the statistical software used. Another outstanding result is that in % 45 of the articles, SPSS was used for quantitative data analysis. Moreover if we ignore the articles where statistical package programs are not reported, this ratio increases to approximately % 89. On the other hand result show us that R software was very few used in the analyzed articles. It is about %0.01 of the total articles published. LISREL, SAS, Amos are also used very few for quantitative data analysis.

Discussion and Conclusions: Results show us that while SPSS is used very common for quantitative data analysis, R is used very rare although it has many advantages. This may be result of that people may think that R is not user friendly because it doesn't have graphical interface. But in fact it is not true. Some studies and trainings should be held to introduce the advantages of R to the researchers.

GİRİŞ

Bilimsel araştırma sürecinin en önemli aşamalarından bir tanesi de elde edilen verileri çözümlenmek ve araştırma problemine ilişkin birtakım bulgulara ulaşmaktır. Elde edilen nicel verilerin çözümlenmesinde ise sıklıkla istatistiksel paket programlarından faydalanılmaktadır. Son yıllarda farklı istatistiksel çözümlenmeler için kullanılan pek çok paket program piyasaya sürülmüştür. Bu paket programlar araştırmacıların nicel verileri çözümlenmeleri sürecinde önemli kolaylık sağlamışlardır. Ancak para ile temin edilen ve her biri farklı çözümlenme teknikleri için özelleştirilmiş paket programların yanı sıra, ücretsiz temin edilen ve pek çok farklı çözümlenme tekniğinin kullanımına olanak tanıyan R yazılımının kullanımı gitgide yaygınlaşmaktadır.

R istatistiksel analiz ve grafik oluşturma sürecinde kullanılan internet üzerinden parasız temin edilebilecek bir yazılımdır. R yazılımının temelini Becker ve Chambers tarafından geliştirilmiş olan "S" dili oluşturmaktadır. R yazılımı daha önce piyasa sunulan ve ücretli bir yazılım olan S-PLUS'ın gelişmiş bir versiyonudur (Er ve Sönmez, 2005; İlk, 2011). R yazılımı yaygın olarak kullanılan bazı istatistik paket programların aksine açık kaynaklıdır (Wenables, Simith ve R Development Core Team, 2002). Para ile temin edilen diğer çoğu paket istatistik programları, yazılımlarının dayandığı kodları geri planda saklayarak kullanıcıya bazı kısa yolların bulunduğu bir ara yüz sunar. Kullanıcı burada ilgili sekmeleri işaretleyerek ve tıklayarak gerekli analizleri gerçekleştirir ancak kullanıcı yazılımın geri planında işleyen ve analizin yapılmasını sağlayan kodları göremez. Açık kaynaklı olan R yazılımı ise bu kodları kullanıcı ile

paylaşır ve dünyanın herhangi bir yerinde herhangi birisinin yazılımın gelişmesine katkı sağlamasına olanak tanır. Bu özellik R yazılımının çok dinamik ve sürekli gelişen bir yapıya sahip olmasını sağlar. Ancak R yazılımını ticari kaygılarla geliştirilen diğer istatistik paket programlarından ayıran temel nokta, geliştirilme amacında yatan felsefesidir. Field, Miles ve Field (2012) R yazılımının ortaya çıkış felsefesini 60'lı yılların meşhur pop rock grubu Beatles'ın ütöpik barış, sevgi ve insanlık vizyonunu ile ilişkilendirmiş ve bu vizyonun istatistik aracılığı ile gerçekleştirildiğini belirtmiştir. Parasız temin edilen ve açık kaynaklı olan R yazılımı, dünyanın farklı yerlerindeki farklı dinlere, dillere, etnik kimliklere vb. sahip bireylerin yazdıkları ve kullandıkları kodlar ile ortak bir platformda buluşmalarını ve birbirlerine destek olmalarını sağlamaktadır. Bu sayede Müslüman bir araştırmacı Musevi bir uzmanın yazdığı kodu kullanarak, herhangi bir ücret ödemediği araştırmadaki gerekli analizleri yapabilirken, benzer bir durum Kübalı ve Birleşik Devletli araştırmacılar için de geçerli olabilmektedir.

Beaujean (2013) R yazılımının 3 önemli avantajını vurgulamıştır. Bunlardan birincisi R yazılımının güçlü bir progamlama diline sahip olması ve bu sayede pek çok nicel analizin hesaplanmasına olanak tanınmasıdır. İkincisi ise kullanıcılara istatistiksel paketleri oluşturmaya imkan tanınması ve bu paketlerin herkes tarafından ulaşılabilir olmasıdır. Üçüncüsü ise daha önce belirtildiği gibi R'in açık kaynaklı bir yazılım olmasıdır.

Dinamik bir yapıya ve sürekli kendini güncelleme olanağına sahip olan R yazılımı temelde bir istatistik yazılım geliştirme ortamıdır. Bu bağlamda diğer istatistiksel paket programların aksine R kullanıcı dostu bir arayüze sahip değildir ve öğrenmesi daha uzun zaman alabilmektedir. Bu durum yeni kullanıcılarda biraz kaygı yaratmaktadır. Ancak kullanıcılar temel mantığını kavradıktan sonra R yazılımının veri çözümleme sürecinde sağladığı özgürlük ve avantajların farkına varabilmektedirler.

R yazılımı parasız temin edilebilmesi, açık kaynaklı olması, araştırmacıların çalışmaları için uygun kendi fonksiyonlarını yazmalarına olanak tanınması, dinamik ve sürekli gelişen bir yapıya sahip olması gibi özellikleri nedeniyle son yıllarda dünyada kullanımı yaygınlaşmıştır. Ancak Türkiye'de R yazılımının kullanımı özellikle sosyal bilimler ve eğitim bilimleri alanlarında çok sınırlıdır. Bunun nedeni R yazılımının yeteri kadar tanınmıyor olması veya araştırmacıların yazılımın kullanımına ilişkin yeterli bilgiye sahip olmamaları olabilir.

Türkiye'de R yazılımının tanıtımına yönelik kaynaklar oldukça sınırlıdır. Er ve Sönmez (2005) R yazılımının kullanımına ilişkin bir makale yayınlamıştır. İlk (2011) ise R yazılımını temel düzeyde anlatan bir kitap kaleme almıştır. Her iki yayında önemli bilgiler içermekle beraber daha çok R yazılımının istatistik alanında yönelik uygulamalarına odaklanmıştır. Bunların yanı sıra R yazılımına yönelik çeşitli kongrelerde sunulmuş sözlü bildirimler bulunmaktadır (Özdemir, Yıldıztepe ve Binar, 2015; Baydoğan, Orbay ve Çetin, 2014). Ancak bu çalışmaların istatistik, biyoistatistik ve mühendislik alanlarında olduğu dikkat çekmektedir. R yazılımının sosyal bilimlerde kullanımına yönelik uygulamaları ve örnekler içeren yayın bulunmamaktadır. Ancak R yazılımının sosyal bilimlerde kullanımının artırılması araştırmacıların nicel verilerin çözümlenmesine ilişkin donanımlarının artması ve araştırmalar için özelleştirilmiş grafiksel gösterimlerin oluşturulabilmesi konularında önemli katkılar sağlayacaktır.

Bu çalışmanın iki temel amacı bulunmaktadır. Bunlardan birincisi R yazılımını tanıtmak ve bazı temel istatistiksel tekniklerin R yazılımında nasıl hesaplanabileceğine ilişkin okuyuculara bilgiler sunmaktır. Araştırmanın ikinci amacı ise Türkiye'de eğitim bilimleri alanında SSCI kapsamında taran dergilerde yayınlanan makalelerde R yazılımının ve diğer istatistiksel yazılımların kullanılma sıklığını yıllara göre betimlemektir. Bu bağlamda çalışma iki bölüme ayrılmıştır. Birinci bölümde R yazılımına ilişkin bazı fonksiyonlar temel düzeyde tanıtılmış, ikinci bölümde ise yukarıda belirtilen alan indekslerinde taranan akademik dergilerde R yazılımının kullanım sıklığı yıllara göre sunulmuştur.

R yazılımı oldukça geniş bir kapsama sahiptir ve üzerinde durulması gereken yüzlerce komut ve fonksiyon bulunmaktadır. Ancak bunların tamamına bu çalışmada değinmek mümkün değildir. Bu nedenle çalışma bazı komut ve fonksiyonlarla sınırlandırılmıştır. Bu bağlamda çalışmada R ile yeni tanışan araştırmacıların öncelikli olarak ihtiyaç duyacağı ve yazılım mantığını temel düzeyde anlamalarına katkı sağlayacak komutlara ve sosyal bilimler - eğitim bilimleri alanlarında sıklıkla kullanılan bazı betimsel ve anlam çıkarıcı istatistiksel tekniklerin

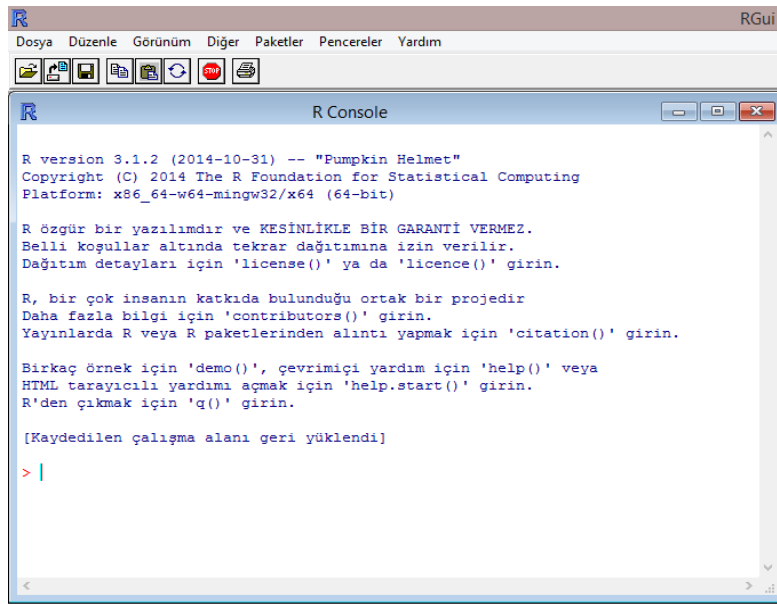
hesaplanmasında kullanılan fonksiyonlara değinilmiştir. Bu bağlamda R yazılımında fonksiyon oluşturmada kullanılan ancak yeni başlayanlar için karmaşık gelebilecek bazı komutlar (loops, conditionals vb.) ve karmaşık desenler, çok değişkenli çözümlere içeren istatistiksel teknikler çalışma kapsamı dışında tutulmuştur. Ancak ilgili araştırmacılar Field, Miles ve Field (2012), Zuur, Ieno ve Smith, (2009) ve Crawley (2007) kaynaklarından daha ayrıntılı bilgilere ulaşabilirler.

Ayrıca çalışmanın betimsel kısmı ise Türkiye’de eğitim bilimleri alanındaki çalışmaların yayınlandığı SSCI kapsamındaki üç dergide 2010-2014 yılları arasında yayınlanan makaleler ile sınırlandırılmıştır. Bu bağlamda çalışmanın birinci bölümünde R yazılımına ilişkin temel bilgiler ikinci bölümünde ise araştırmanın betimsel kısmına yönelik yöntem ve bulgular sunulmuştur.

R Yazılımını Temin Etmek

R yazılımını temin etmek için (<http://www.R-project.org/>) internet sayfasını ziyaret edip sayfada bulunan “download R” butonuna tıklamanız gerekmektedir. Daha sonra açılan sayfadan “Turkey” başlığı altında bulunan linki tıklamanız gerekmektedir. Son aşama olarak da açılan sayfada bilgisayarınızın işletim sisteminin yazılı olduğu (Windows, Linux, Mac) linki tıklayarak gerekli yönlendirmeleri de takip ederek indirme işlemini başlatabilirsiniz. İndirme işlemi tamamlandıktan sonra R yazılımının kısa yoluna tıklayarak başlatabilirsiniz. Yazılım temel ara yüzü Şekil 1’de verilmiştir.

Şekil 1.



```
R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R özgür bir yazılımdır ve KESİNLİKLE BİR GARANTİ VERMEZ.
Belli koşullar altında tekrar dağıtımına izin verilir.
Dağıtım detayları için 'license()' ya da 'licence()' girin.

R, bir çok insanın katkıda bulunduğu ortak bir projedir
Daha fazla bilgi için 'contributors()' girin.
Yayınlarda R veya R paketlerinden alıntı yapmak için 'citation()' girin.

Birkaç örnek için 'demo()', çevrimiçi yardım için 'help()' veya
HTML tarayıcılı yardımı açmak için 'help.start()' girin.
R'den çıkmak için 'q()' girin.

[Kaydedilen çalışma alanı geri yüklendi]

> |
```

R yazılımının kullanıma yönelik bazı temel komut ve fonksiyonlara ilişkin bilgi vermeden önce bazı kavramlara değinmekte fayda vardır. Daha öncede belirttiği gibi dünyanın pek çok yerindeki uzmanlar belirli istatistiksel analizler için kodlar yazarak paketler oluşturmakta ve böylece R yazılımının gelişimine katkı sağlamaktadır. Yazılan bu istatistik paketleri ve aynı zamanda yazılımın kendisi de CRAN (Comprehensive R Archive Network) adı verilen merkezi konumlarda muhafaza edilmektedir. Başka bir ifade ile CRAN bir R kullanıcısının yapmak istediği belirli bir istatistiksel analize yönelik hazırlanmış paketleri indirdiği sanal bir ortamdır. Bu sanal ortamın, dünyanın pek çok yerinden kişilerin girmesi neticesinde yavaşlama veya internet korsanları tarafından saldırıya uğratılıp çökertilme olasılığı söz konusudur. Bu nedenle tek bir CRAN merkezine sahip olmaktan ziyade, bu merkezin yansuları (kopyaları) dünyanın pek çok bölgesine dağıtılmıştır. R yazılımında “CRAN yansı” terimi merkez CRAN ortamının birebir kopyası anlamında kullanılmaktadır. Dünyanın pek çok bölgesinde CRAN yansuları bulunmaktadır. Bir R kullanıcısı kullanacağı istatistiksel paketi ya da programın kendisini

indirmek için coğrafi olarak kendisine en yakın olan CRAN yansısını seçme durumundadır (Zuur,Ieno ve Messters, 2009).Örneğin Türkiye’de bulunan bir R kullanıcısı “Turkey” yansısını seçmelidir. Bazı ülkelerde ise birden fazla CRAN yansısı bulunabilmektedir. Eğer bulunduğunuz ülke için oluşturulmuş bir CRAN yansısı yok ise coğrafi olarak en yakın yansının seçilmesi gerekmektedir.

R yazılımı diğer istatistiksel yazılımların aksine grafiksel bir ara yüzle değil gerekli kod komutların veya fonksiyonların yazılması ile çalışır. Diğer bir ifade ile yazılıma ne yapması gerektiği onun dilinde anlatılır. Bu bağlamda R yazılımını kullanmak temelde iki beceriyi gerektirmektedir. İdeal bir R kullanıcısı hesaplanmak istenen istatistiğin temel bileşenlerini ve bu istatistiğin nasıl hesaplanması gerektiğini yazılıma öğretmek için R dilini bilmesi gerekmektedir. Bir kullanıcı bu becerilere sahipse R yazılımı o kişinin kurguladığı tüm istatistiksel analizleri yapabilme gücüne sahip olacaktır. Ancak herkesin bu becerilere yeterli düzeyde sahip olması mümkün değildir. Bazı kullanıcılar istatistiği yeterli düzeyde bilirken R diline hakim olamayabilir ya da R dilini bilirken yeterli istatistiksel bilgiye sahip olamayabilir. Bu nedenle dünyanın dört bir yanında bu becerilere sahip uzmanlar tarafından belirli istatistiksel analizlerin hesaplanması için hazırlanan paketler bulunmaktadır. Bu paketler CRAN adı verilen ortamlarda muhafaza edilmekte ve kullanıcılar tarafından indirilebilmektedir. Ancak yine de bazı temel kodların yazılmasına ihtiyaç duyulmaktadır. R yazılımının diğer paket programlardaki gibi kullanıcı dostu olan ve açıklayıcı bilgiler içeren grafiksel ara yüzünün olmaması ve açılan bir pencereye kod ve komut ve fonksiyonların yazılması gerekliliği kullanıcılara başlangıçta biraz zor gözükmeştir. Ancak zamanla R yazılımı kullanıcılara kolay gelmekte ve çok önemli avantajlar sunmaktadır.

R Yazılımında kullanılan Kodların Temel Yapısı

R yazılımında “>” sembolü satır başını ifade eder ve kod yazılan her yeni satırın başında bulunur. R yazılımında komutlar genelde nesnelere ve fonksiyonlar olmak üzere iki temel boyuttan oluşur. Bir satırda bu iki boyut “<-” sembolü ile birbirinden ayrılır. Bu sembolün sol tarafında kalan kısım **nesne** iken sağ tarafında kalan kısım ise **fonksiyon** boyutudur (nesne <- fonksiyon). “<-” sembolü R yazılımına belirtilen nesne belirtilen fonksiyon tarafından oluşturulacaktır komutunu vermektedir. R yazılımında oluşturulan herhangi bir şey “nesne” olarak ele alınır (Knell, 2014). Nesne bir değişken veya bir istatistiksel model olabilir. Fonksiyonlar ise nesneyi oluşturması için R yazılımına verdiğiniz kodlardır. Bir örnek verecek olursak;

```
> puan <-c(10,20,30,40,50)
```

şeklinde bir komut oluşturduğumuzda yazılıma “<-” sembolünün sağ tarafında bulunan değerlerden “puan” değişkenini oluşturmasını istemiş oluruz. Burada puan değişkeni nesne iken sembolün sağ tarafında bulunan kısım ise fonksiyondur. Burada kullanılan “**concatenate**” fonksiyonu kısaca “**c()**” yazılıma değişkenleri bir grup altında topla komutunu vermektedir. Benzer bir şekilde nitel değerlerden oluşan bir değişken oluşturmak istendiğinde de c() fonksiyonu kullanır ancak bu sefer değişken verileri nitel olduğu için o değerlerin tırnak içerisinde (“ kız” vb.) girilmesi gerekmektedir. Şu şekilde örnek bir komut sunulabilir;

```
> cinsiyet<-c("kız","erkek","kız","erkek")
```

Bu sayede R yazılımı bir değişkenin nitel mi yoksa nicel mi olduğuna karar verir. Bu aşamadan sonra her “puan” veya “cinsiyet” nesnesi girildiği ve “enter” tuşuna basıldığı zaman yazılım ilgili değişkenleri getirecektir. Bu süreçte dikkat edilmesi gereken önemli bir nokta R yazılımının büyük küçük harfe duyarlı olmasıdır. R yazılımı biri büyük diğeri küçük harfle yazılmış aynı iki komutu tamamen farklı algılayacaktır. R yazılımında eğer bir satırda yazılan fonksiyon bir şekilde tamamlanmamış veya eksik bırakılmış ise kullanıcı “enter” tuşuna bastığında R bir sonraki satıra “>” sembolü yerine “+” sembolü ile başlar. Bu şekilde R kullanıcıya komut veya fonksiyonu yazarken bir boyutunu eksik bıraktığı uyarısını yapar ve

aşağıdaki satıra eksik kısmı yazması gerektiğini belirtir. Bu durumu içeren örnek aşağıda verilmiştir.

```
> örnek<-c(10,20,30,40  
+ |
```

Bu örnekte kullanıcı açtığı parantezi kapatmayı unuttuğu için bir sonraki satır “+” sembolü ile başlamıştır. R yazılımına ilişkin bazı temel fonksiyonlar ileriki bölümlerde sunulacaktır.

Diğer Paket Programlardan R yazılımına Veri Aktarımı

R yazılında önemli boyutlardan birisi de diğer paket programlarda bulunan verilerin R platformuna aktarım işlemidir. R yazılımı içerisinde de verilerin oluşturulması mümkündür ancak yaygın olarak SPSS, Excel gibi paket programlara girilmiş verilerin aktarılmasının gerektiği durumlar da olmaktadır. Temel olarak piyasada bulunan mevcut tüm paket programlardan R yazılımına veri aktarımı yapılabilmektedir. Ancak bu süreçte verinin nasıl bir programdan aktarıldığının ve verinin yapısının nasıl olduğunun R yazılımına tanıtılması gerekmektedir. Bunun için de bazı kısa fonksiyonların yazılmasının gerekliliği vardır. Ancak bir veri R yazılımına aktarılıp daha sonra da bir kez kaydedilirse, ilgili veriler R yazılımı bilgisayardan silinmediği müddetçe tek bir tuşla ulaşılabilir konumda olacaktır. Bu bakımdan R yazılımı diğer paket programlara kıyasla daha pratik bir yapıya sahiptir. Bu çalışmada yaygın olarak kullanılan SPSS ve Excel dosyalarından veri aktarımının nasıl yapıldığına yönelik örnek fonksiyonlar sunulmuştur.

R yazılımına en kolay veri aktarımının yapıldığı program ve dosya türlerinden birisi CSV uzantılı Excel dosyalarıdır. Bir Excel dosyasının uzantısı CSV olarak değiştirildikten sonra R yazılımına aktarımı daha sorunsuz olmaktadır. Veri aktarımında öncelikli olarak, aktarılacak veri dosyasına verilecek adın belirlenmesi gerekmektedir ki bu kısım yazılacak komutun nesne kısmını oluşturacaktır. Daha sonra **read.table()** fonksiyonu ile aktarılacak dosyanın bilgisayardaki konumu ve verinin bazı özellikleri girilir. Örnek bir komut aşağıda verilmiştir. Yukarıdaki komut incelendiğinde nesne bölümünün “<-” sembolünün sol kısmında bulunan

```
> örnek<-read.table("C:\\Users\\Bilgisayarım\\Desktop\\TUTUM.csv",header=T,sep=";")
```

“örnek” kısmı olduğu görülmektedir. “<-” sembolünün sağ tarafında kalan kısım ise komutun fonksiyon bölümünü oluşturmaktadır. Bu kısımda öncelikli olarak read.table fonksiyonu girildikten sonra parantez içerisinde dosyanın bilgisayarda bulunduğu konumu yazılır. Dosyaların bilgisayarda hangi konumda saklanacağı “*working directory*” (çalışma dizini) ile belirlenir. R yazılımında yazdığınız (oluşturduğunuz) dosyaları belirli bir yere kaydetmek istediğiniz de veya belirli bir yerde bulunan dosyaları R yazılımına aktarmak istediğimizde bir çalışma dizini oluşturulması yerinde olacaktır. Bu süreçte getwd() ve setwd komutları kullanılır. getwd() komutu ile mevcut dizin görülebilirken yeni bir dizin oluşturmak istendiğinde setwd() komutu kullanılabilir. Dosyanın bilgisayardaki konumu girildikten sonra belirtilen **header=T** kısmı ile verinin alınacağı Excel dosyasında ilk satırda başlıklar (değişken adları) olduğu ve verilerin ondan sonraki satırdan başladığı ifade edilir. “T” İngilizce True(doğru) anlamındadır. Eğer **header=F** yani İngilizce False(yanlış) olarak yazılırsa R yazılımı ilk satırı veri satırı olarak algılayacaktır (Crawley, 2007). Akabinde belirtilen “sep” fonksiyonu verilerin birbirinden nasıl ayrıldığı ile ilgilidir. CSV (comma separated variable) virgülle ayrılmış değişken anlamına gelir. Ancak noktalı virgül(;) ile ayrıldığı durumlar da söz konusudur. Bu noktada CSV uzantılı Excel dosyası farede bulunan sağ klik ile tıklanıp “notepad” ile açıldığında dosyada bulunan verilerin virgülle mi yoksa noktalı virgülle mi ayrıldığı kolaylıkla belirlenebilir. Bu bağlamda dosya noktalı virgül ile ayrılmış ise “sep” fonksiyonu sonrasında tırnak içinde “;” yazarız. Bu sayede R yazılımı dosyayı tanır ve hatasız bir şekilde açılmasını sağlar.

Eğer ilgili dosyaların uzantılarını yazmak zor geliyorsa **read.csv(file.choose())** fonksiyonu kullanılır. Bu fonksiyon çalıştırıldığında aynı diğer paket programlarda olduğu gibi

bir pencere açılır ve o pencereden ilgili dosyayı seçerek R ortamına aktarabilirsiniz. Aşağıda örnek bir komut verilmiştir.

```
> örnek<-read.csv(file.choose(),header=T,sep=";")
```

R yazılımına SPSS üzerinden veri aktarıldığı durumlar için eğer hali hazırda yüklü değil ise “foreign”adlı paketinin yüklenmesi gerekmektedir. Paketlerin nasıl indirileceğine ileriki bölümlerde değinilecektir. Bir SPSS dosyasında R ortamına veri aktarımında da benzer bir mantık söz konusudur. Ancak Excel dosyalarındakinden farklı bazı fonksiyonlar bulunmaktadır. Örnek bir komut aşağıda verilmiştir;

Yukarıda verilen örnek komut incelendiğinde Excel dosyaları için kullanılan **read.table** veya

```
> örnek1<-read.spss("C:\\Users\\Bilgisayarım\\Desktop\\R data\\AB.sav",use.value.labels=TRUE,to.data.frame=TRUE)
```

read.csv fonksiyonu yerine **read.spss** fonksiyonun kullanıldığı dikkat çekmektedir. Akabinde yine dosyanın bilgisayardaki konumu girilmiştir. Bu boyuttan sonra Excel dosyasından veri aktarırken kullanılmayan iki adet fonksiyon bulunmaktadır. Bunlardan birincisi “**use.value.labels=TRUE**”fonksiyonudur. Bu fonksiyon “**use.value.labels=FALSE**” şeklinde de yazılabilir. Bir SPSS dosyasına nicel veriler girmekle beraber sınıflamalı ya da sıralamalı düzeyde olan ancak nicel olarak girilmiş verilere etiket atanabilmektedir. Örneğin bir veri grubunda 1 değeri “erkek” 2 değeri ise “kadın” olarak etiketlenebilmektedir. Eğer bu fonksiyon“**use.value.labels=TRUE**” şeklinde yazılırsa R nicel değerler yerine o değerlere karşılık gelen etiketleri veri dizisinde gösterecektir. “**use.value.labels=FALSE**” şeklinde yazılırsa ise nicel değerleri veri dizisinde gösterecektir. Diğer fonksiyon olan **to.data.frame** ise mutlaka “**to.data.frame=TRUE**” şeklinde kullanılmalıdır. Aksi takdirde R yazılımı verileri çok karışık ve her değişkeni farklı bir satır üzerinde gösterecek şekilde aktaracaktır. Hangi programdan olursa olsun R yazılımına veri aktarımında, kaynak veri dosyasının bilgisayar içerisindeki konumu, o dosyanın kısa yolunun fare ile tutulup R yazılımının içerisine sürüklenmesi ile ekranda kolayca görülebilir. İlgili kısım buradan kopyalanarak kolaylıkla kullanılabilir. Böyle bir durumda R hata uyarısı verecektir ancak buradaki temel amaç dosya konumunu belirlemek olduğu için bu hata uyarısı göz ardı edilmelidir.

Bir SPSS dosyasını **file.choose()** fonksiyonu ile R ortamına aktarmak istendiğinde aşağıdaki komut da kullanılabilir.

```
> örnek<-read.spss(file.choose(),use.value.labels=F,to.data.frame=T)
```

Bu komut sonrasında diğer paket programlarda olduğu gibi ilgili dosyayı seçmemiz R ortamına aktarılması için yeterli olacaktır. Bunların yanı sıra R yazılımına veri aktarma sürecinde kullanılan başka fonksiyonlar da yer almaktadır. Örneğin **read.rwf()** sabit formatlı text dosyalarını R yazılımına aktarma sürecinde kullanılır. **read.dta()** komutundan ise Stata (.dta uzantılı) veri dosyalarını aktarmada faydalanılır.

İstatistik Paketlerinin İndirilmesi

R yazılımı bilgisayara indirildiğinde bazı temel paketleri içermektedir. Ancak hesaplanmak istenen spesifik bir istatistiksel analiz için bazı paketlerin indirilmesi gerekmektedir. Örneğin bir psikometrist madde tepki kuramına (MTK) dayalı madde parametrelerinin tahmini için “**sirt**” paketini veya doğrulayıcı faktör analizi hesaplamak isteyen bir araştırmacı ise “**sem**” paketini bilgisayarına indirmelidir. Gerekli istatistiksel paketin indirilmesi için **install.packages (“paketin adı”)** fonksiyonun yazılması yeterlidir (Crawley, 2007; Zuur,Ieno ve Meesters, 2009). Bir örnek verecek olursak MTK’ya dayalı bazı parametre tahminlerinin yapılması gerektiğinde aşağıdaki fonksiyon kullanılmalıdır.

```
> install.packages("sirt")
```

Burada dikkat edilmesi gereken nokta ilgili paket isminin hem parantez hem de tırnak içinde yazılı olmasıdır. İlgili fonksiyon yazılıp enter tuşuna basıldıktan sonra ekranda belirlenen ara yüzde ilgili paketi hangi CRAN yansısından indireceğinizi seçmeniz istenir. Daha öncede belirtildiği gibi bu aşamada coğrafi olarak bulunulan konuma en yakın olan CRAN yansısının seçilmesi gerekmektedir. Şu an kullanıma hazır ve çeşitli istatistiksel analizler için geliştirilmiş 5000'den fazla paket bulunmaktadır ve bu sayı sürekli artmaktadır. İlgili paket bilgisayara indirildikten sonra R yazılımının kütüphanesinde yerini alacaktır. Bu işlemlerin yapılabilmesi için internet bağlantısının bulunması gereklidir. Analiz yapılmadan önce **library(paketin adı)** fonksiyonu yazılarak yüklenmiş olan paketin çağrılması gerekmektedir.

```
> library(sirt)
```

Burada ilgili paketin ismi tırnak içinde değil sadece parantez içinde yazılmalıdır. R yazılımı içerisinde farklı uzmanlar tarafından yazılmış binlerce paketi içerir. Uzmanlar bu istatistik paketleri içerisinde birbirlerinden habersiz olarak aynı ada sahip fonksiyonlar farklı işlevler için kullanabilmektedir. Örneğin hem “**car**” hem de “**Hmisc**” paketleri “**recode**” fonksiyonuna sahiptir. Eğer iki pakette indirilmiş ise kullanıcı “**recode**” fonksiyonunu yazdığı anda yazılım hangi pakete ait olan fonksiyonun kastedildiğini algılayamaz veya hatalı bir paketi seçebilir (Field, Miles ve Field, 2012). Bu nedenle yukarıda örneği verilen **library()** fonksiyonu ile ilgili hangi paketin kullanılacağı bilgisi yazılıma verilir. Ayrıca **update.package()** fonksiyonu ile daha önceden bilgisayarınıza indirmiş olduğunuz bir paketi güncelleyebilirsiniz.

R Yazılımı ve Mobil Uygulamalar

R yazılımı diğer istatistik paket programlarının aksine çok güçlü bir bilgisayar donanımı gerektirmez ve akıllı telefonlarda veya tablet bilgisayarlarda dahi kolaylıkla kullanılabilir. Hem IOS hem de Android platformlarında “R Console” uygulaması indirilerek kolaylıkla kullanılabilir. “R Console” uygulaması iki versiyona sahiptir. Parasız temin edilebilen “R Console Free” uygulaması belli özellikler bağlamında sınırlandırılmıştır. “R Console Premium” uygulaması ise cüzzü bir ücrete sahip olup bilgisayarlarda kullanılan R yazılımının tüm özelliklerine sahiptir. “R Console” uygulamasında dışardan veri aktarımının yapılabilmesi için “dropbox” veya “skydrive” benzeri sanal depolama alanlarından faydalanılmaktadır. Bu depolama alanlarında bulunan veri dosyaları internet bağlantısı olan bir akıllı telefon ya da tablet bilgisayarda bulunan “R Console” uygulamasına kolaylıkla aktarılabilir. Bu sayede bir araştırmacının metroda giderken veya parkta otururken akıllı telefonunda bulunan “R Console” uygulaması ile araştırması için gerekli analizleri yapılması mümkün olmaktadır. “R Console” uygulamasında yukarıda belirtilen platformlardan ulaşılamadığı durumlarda arama motorlarından tarama yapılarak da akıllı telefon ve tablet bilgisayarlara indirilebilmektedir. Ancak bunun için cihazın güvenlik ayarlarında küçük değişikliklerin yapılması gerekebilir.

R Yazılımına İlişkin Temel Fonksiyonlar

Çalışmada özellikle sosyal bilimler alanında araştırmacılara katkı sağlayacağı düşünülen bazı temel fonksiyonlar “Betimsel İstatistikler”, “Temel Varsayımların Test Edilmesi” ve “Anlam Çıkarıcı İstatistikler” başlıkları altında sunulmuştur.

Betimsel İstatistikler

summary(dosya adı) : Bu fonksiyon ile belirtilen dosyada bulunan tüm değişkenler için bazı betimsel istatistikler elde edilir. Belirtilen fonksiyon yazılıp “enter” tuşuna basıldığında R dosyada bulunan tüm değişkenlere ilişkin en küçük ve en büyük değerleri, birinci ve üçüncü çeyreklikleri, ortanca ve aritmetik ortalama değerlerini verir. Aşağıda daha önce örneklendirilen “Tutum” dosyasına ilişkin fonksiyon ve çıktılar sunulmuştur.


```
> summary(Tutum)
      ÖNTEST      SONTEST      İZLEME
Min.   : 9.00   Min.   : 45.0   Min.   : 81.0
1st Qu.: 16.00  1st Qu.: 50.0   1st Qu.: 256.0
Median : 20.00  Median : 65.0   Median : 400.0
Mean   : 35.56  Mean   :108.9   Mean   : 806.7
3rd Qu.: 30.00  3rd Qu.: 70.0   3rd Qu.: 900.0
Max.   :160.00  Max.   :490.0   Max.   :3630.0
```

mean(dosyanın adı \$ dosyada bulunan değişkenin adı): Mean fonksiyonu ile ilgili değişkene ilişkin aritmetik ortalama değeri hesaplanır. Ancak dosyada yer alan hangi değişken için ilgili analizin hesaplanmak istendiğinin belirtilmesi gerekmektedir. Bunun için "\$" sembolü kullanılır. "Mean" fonksiyonundan sonra parantez içerisinde dosyanın adı yazıldıktan sonra "\$" sembolü yazılarak değişkenin adı yazılır. "\$" sembolü bir dosya içerisindeki ilgili değişkenin seçilmesi sürecinde kullanılır. Örneğin;

```
> mean(Tutum$ÖNTEST)
[1] 35.55556
```

funksiyonu ile "Tutum" dosyasında bulunan "ÖNTEST" değişkenine ilişkin aritmetik ortalama değeri hesaplanır. Aynı işlem **sd (dosyanın adı \$ dosyada bulunan değişkenin adı)** fonksiyon ile yapılarak ilgili değişkene ilişkin standart sapma değeri hesaplanabilir. Örnek bir fonksiyon aşağıda verilmiştir.

```
> sd(Tutum$SONTEST)
[1] 143.3697
```

describe(dosya adı): Bu fonksiyon ile dosyada bulunan tüm değişkenlere ilişkin aritmetik ortalama, ortanca, standart sapma, basıklık ve çarpıklık katsayıları gibi bazı temel betimsel istatistikler üretilir. Ancak bu fonksiyonun kullanılması için "**psych**" paketinin indirilmesi gerekmektedir. Örnek bir fonksiyon ve çıktısı aşağıda verilmiştir.

```
> describe(Tutum)
      vars n  mean      sd median trimmed  mad min  max range skew kurtosis  se
ÖNTEST   1 9 35.56  47.24   20  35.56  11.86   9  160   151 1.98    2.38  15.75
SONTEST   2 9 108.89 143.37   65 108.89  22.24  45  490   445 2.05    2.56  47.79
İZLEME    3 9 806.67 1100.51  400 806.67 379.55  81 3630  3549 1.79    1.86 366.84
```

Temel Varsayımların Test Edilmesi

Betimsel istatistiklerin aksine anlam çıkarıcı istatistiksel analizler hesaplanmadan önce bazı varsayımların veri üzerinde test edilmesi gerekliliği söz konusudur. Bu çalışmada yalnızca tek değişkenli normallik ve varyansların homojenliği varsayımlarına yönelik R fonksiyonları sunulmuştur.

Tek değişkenli Normallik Varsayımı

shapiro.test(dosya adı \$ değişken adı):Bu fonksiyon ile ilgili değişkene yönelik dağılımın normalliği hipotez testi ile sınanır. Elde edilen sonuç istatistiksel olarak anlamlı değil ise ($p > 0,05$) dağılımın normallik varsayımını karşıladığı düşünülür (Field, 2009). Aşağıda "Tutum" dosyasında bulunan "İZLEME" değişkeni için hesaplanmış örnek bir fonksiyon ve çıktı sunulmuştur.

```
> shapiro.test(Tutum$İZLEME)

      Shapiro-Wilk normality test

data:  Tutum$İZLEME
W = 0.6402, p-value = 0.0002834
```

by(dosya adı \$ değişken adı, dosya adı \$ grup değişkeninin adı, shapiro.test): Bazı durumlarda örneğin bağımsız örneklem için t testi hesaplanırken normallik varsayımının alt gruplar için ayrı ayrı test edilmesi gerekmektedir. Böyle bir durumda “by” fonksiyonundan faydalanılır. “Tutum” dosyasında bulunan “İZLEME” değişkeninin cinsiyet alt grubuna göre normallik testi sonuçları aşağıda sunulmuştur.

```
> by(Tutum$İZLEME, Tutum$CİNSİYET, shapiro.test)
Tutum$CİNSİYET: erkek
```

```
Shapiro-Wilk normality test
```

```
data: dd[x, ]
W = 0.7704, p-value = 0.0589
```

```
-----
Tutum$CİNSİYET: kız
```

```
Shapiro-Wilk normality test
```

```
data: dd[x, ]
W = 0.9229, p-value = 0.5489
```

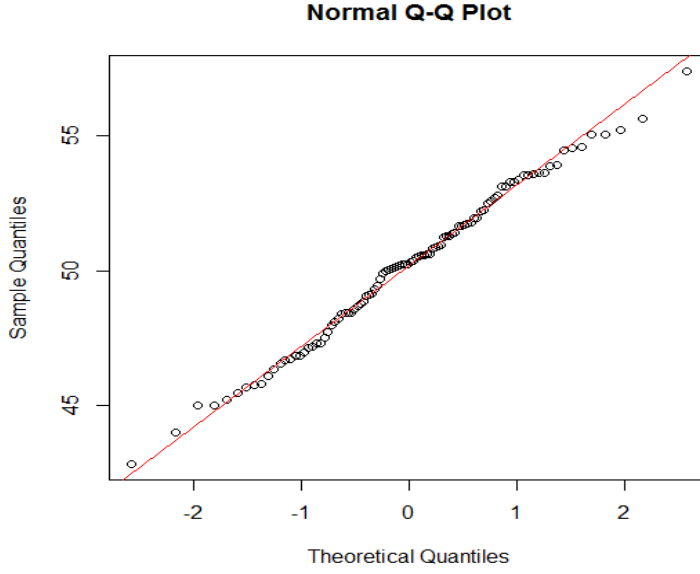
Tek değişkenli normallik varsayımının test edilmesinde kullanılan bir diğer hipotez testi ise Kolmogrov Simirnov testidir. Kolmogrov Simirnov testi grup büyüklüğünün 50’den büyük olması durumunda Shapiro Wilk testinin yerine kullanılır. Ancak R yazılımı içerisinde yer alan ks.test() fonksiyonu bir dağılımın normallikten sapma düzeyinin belirlenmesinden ziyade iki farklı dağılımı karşılaştırmak amacı ile kurgulanmıştır. Bu nedenle ks.test() fonksiyonu kullanılırken normalden sapma düzeyi belirlenecek dağılımın dosya adının yanı sıra “pnorm” komutu da belirtilir. “pnorm” komutu ortalaması 0 standart sapması 1 olan kusursuz normal dağılımı temsil eder. Bu sayede fonksiyon normallik varsayımı test edilecek değişkenin dağılımı ile ortalaması 0 standart sapması 1 olan bir dağılımı karşılaştırmış olur. Bu bağlamda fonksiyon şöyle oluşturulabilir. **ks.test (değişken adı, pnorm)**

Tek değişkenli normallik varsayımı hipotez testlerinin yanı sıra bazı grafikler incelenerek de test edilebilir ki örneklem büyük olduğu durumlarda (N>200) grafiklerin incelenmesi ve bazı betimsel analiz sonuçlarının yorumlanması önerilir (Field, Miles ve Field, 2012). Grafikler arasında en sık kullanılanı ise Q-Q plot grafikleridir.

qqnorm(dosya adı \$ değişken adı ; qqline(dosya adı \$ değişken adı, col="red")):Bu fonksiyon Q-Q plot grafiği oluşturmada kullanılan iki fonksiyonun birleştirilmiş halidir. **qqnorm**fonksiyonu ilgili değişkendeki verilerin dağılımını, **qqline** fonksiyon ise dağılım sahip olması beklenen normal dağılımı gösteren çizgiyi oluşturur. Verilerin dağılımının belirtilen çizgiden ne düzeyde sapma gösterdiğine dayalı olarak dağılımın normalliğine ilişkin bir karara varılır. **qqline** fonksiyonu sonunda grafiğin daha net görülebilmesi için normal dağılımı simgeleyen çizginin rengi de belirtilebilir.

Aşağıda “Tutum” dosyasında bulunan “Başarı” değişkeni için örnek bir fonksiyon ve çıktı sunulmuştur.

```
> qqnorm(Tutum$Başarı); qqline(Tutum$Başarı,col="red")
```



stat.desc(dosya_adi,basic=F, norm=T): Bu fonksiyon tek deęişkenli normallięinin test edilmesinde kullanılan bazı istatistiklerin hesaplanması için kullanılır. Bu fonksiyonun çalışması için “**pastecs**” paketinin indirilmesi gerekmektedir. Fonksiyonda bulunan “**basic=F**” bölümü bazı temel analizlerin hesaplanmasının istenmedięi anlamına gelir. Bu kısım “**basic=T**” şeklinde de yazılabilir ancak asıl odak noktası normallik varsayımının test edilmesi olduęu için bu kısmın “**F**” olarak yazılması daha pratik olacaktır. “**norm=T**” ise puanların dağılımına (normallik varsayımına) yönelik gerekli analizlerin istendięi anlamına gelmektedir. Aşağıda “örnek” isimli bir dosya için fonksiyon ve çıktıları sunulmuştur.

```
> stat.desc(örnek,basic=F,norm=T)
              başarı    yetenek
median      9.6734055  9.96655569
mean       10.2555847  9.90023247
SE.mean     0.5623731  0.50180950
CI.mean.0.95 1.1158703  0.99569892
var         31.6263549  25.18127777
std.dev     5.6237314  5.01809503
coef.var    0.5483580  0.50686638
skewness    0.1996297 -0.03909600
skew.2SE    0.4135178 -0.08098441
kurtosis    -0.3610523 -0.35378778
kurt.2SE    -0.3774083 -0.36981472
normtest.W  0.9874726  0.99028368
normtest.p  0.4701676  0.68753401
> |
```

Bir dağılıma ilişkin tek deęişkenli normallik varsayımı test edilirken çarpıklık ve basıklık katsayılarının kendi standart hatalarına bölümü de dikkate alınan bir deęerdir. Bu deęerin 1,96 yaklaşık olarak 2’den büyük olması 0,05 düzeyinde sonucun anlamlı olduęu yani dağılımın normallik varsayımını karşılamadıęı şeklinde yorumlanır. Tabloda verilen skew.2SE ve kurt.2SE deęerleri çarpıklık ve basıklık katsayısının kendi standart hatalarının 2 ile çarpımına bölünmesi ile elde edilen deęerlerdir. Bu deęer için de benzer bir mantıkla 1’den büyük olması sonucun 0,05 düzeyinde anlamlı olduęu, 1’den küçük olması ise anlamlı olmadığı yani dağılımın normal olduęu şeklinde yorumlanır. Yine çıktının sonunda bulunan “normtest.p” deęeri ise daha önce verilen **shapiro.test** fonksiyonu elde edilen hipotez testi sonucuna eşittir.

Varyansların Homojenliği

Varsayanların homojenliği normallik varsayımının yanı sıra anlam çıkarıcı pek çok istatistik için test edilmesi gereken bir ön koşuldur. Varyansların homojenliği varsayımı yaygın olarak Levene testi hesaplanarak kontrol edilir (Tabachnickand Fidel, 1996; Palland, 2002).

leveneTest (Dosya adı \$ çıktı değişkeni, Dosya adı \$ grup değişkeni): Levene testinin hesaplanmasında kullanılan bu fonksiyonunu çalıştırmak için "car" paketinin indirilmesi gerekmektedir. Levene testinde de 0,05'ten büyük olan p değerleri varyansların homojenliği varsayımının karşılandığı anlamına gelir. Aşağıda örnek bir veri için fonksiyonu ve analiz çıktıları sunulmuştur.

```
> leveneTest(Tutum$ÖNTEST, Tutum$CİNSİYET)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  1.2499 0.3005
      7
```

Anlam Çıkarıcı İstatistikler

Bu bölümde sosyal bilimlerde sık kullanılan bazı anlam çıkarıcı istatistiksel tekniklerin hesaplanmasına yönelik bazı R fonksiyonları ve çıktıları sunulmuştur.

Bağımsız Örneklem için T testi

Bağımsız örneklem için T-testi ilişkisiz örneklem ortalamaları arasındaki farkın manidarlığını test etmek için kullanılır. Bağımlı değişkene ait ölçümlerin en az eşit aralıklı olması bağımsız değişkenin ise 2 kategorili süreksiz olması gerekir. Bağımlı değişkene ait ölçümlerin ise her iki grup için de normal dağılım göstermesi ve varyanslarının homojen olması beklenir. Ancak güçlü bir parametrik test olan bağımsız örneklem için t testi varyansların homojenliği varsayımının karşılanmadığı durumlarda da hesaplanabilir (Büyüköztürk, 2010).

t.test(bağımsız değişken ~ bağımlı değişken data= dosya adı): Bu fonksiyon bağımsız değişkenler için t testinin hesaplanmasında yaygın olarak kullanılmaktadır (Crawley, 2007). Ancak analize geçmeden önceki bölümlerde belirtilen kodların kullanılarak gerekli varsayımların test edilmesi gerekmektedir. Aşağıda örnek bir kod ve çıktı verilmiştir. R yazılımı "Welch Two sample Test" adında bir test hesaplamaktadır. Ancak bu analiz bağımsız örneklem t testi ile aynı hesaplama adımlarına sahiptir ve muadili konumdadır.

```
> t.test(SONTEST~CİNSİYET, data=Tutum)

Welch Two Sample t-test

data:  SONTEST by CİNSİYET
t = 1.0036, df = 3.021, p-value = 0.389
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -232.5466  448.0466
sample estimates:
mean in group erkek    mean in group kız
      168.75             61.00
```

Bağımlı Örneklem için T testi

İlişkili iki örneklem ortalaması arasındaki farkın manidar olup olmadığını test etmek için bağımlı (tekrarlı) örneklem için t testi hesaplanır. Bu test bağımlı ve bağımsız değişkenin en az eşit aralık ölçeğinde ve ölçümlerin dağılımlarının normallik varyansların homojenliği varsayımlarını karşılamış olmasını ve gerektirir. Ancak bağımsız örneklem için t testinde olduğu gibi varyansların homojenliği varsayımının ihlali durumlarında analize devam edilebilir.

t.test(dosya adı \$ birinci değişken adı, dosya adı \$ ikinci değişken adı, paired= T)

Bu fonksiyon bağımlı örneklem için t testinin hesaplanmasında yaygın olarak kullanılmaktadır. Fonksiyonda bulunan “**paired=T**” ifadesi iki örneklemin ilişkili olduğu bilgisini yazılıma sunar. Aşağıda örnek bir fonksiyon ve çıktı sunulmuştur.

```
> t.test(Tutum$SONTEST,Tutum$ÖNTEST,paired=T)
```

```
Paired t-test
```

```
data: Tutum$SONTEST and Tutum$ÖNTEST
t = 2.2827, df = 8, p-value = 0.05185
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.7494796 147.4161463
sample estimates:
mean of the differences
      73.33333
```

NOT: Bu bölümde belirtilen tüm fonksiyonlar bir nesne oluşturularak ta yazılabilir. Bir fonksiyon araştırmacının tercihinin bağlı olarak önceki bölümlerde belirtildiği gibi kendi belirleyeceği bir nesne ile ilişkilendirilebilir. Örneğin aşağıda verilen bağımlı örneklem için t testi hesaplamada kullanılan fonksiyon “örneklem 1 “ nesnesi ile ilişkilendirilebilir. Yani örneklem 1 nesnesi “<-“ sembolünün sağ tarafında bulunan fonksiyon ile oluşturulmuş olur. Bu sayede araştırmacı R ortamında “örneklem1” yazdığı her zaman tekrar aynı fonksiyonu yazmadan sadece oluşturduğu nesneyi yazarak ilgili analiz çıktılarını ekranda görebilecektir.

```
> t.test(SONTEST~CİNSİYET,data=Tutum)
```

```
> örneklem1<-t.test(SONTEST~CİNSİYET,data=Tutum)
```

```
> örneklem1
```

Tek Yönlü Varyans Analizi (Tek Yönlü ANOVA)

Tek yönlü varyans analizi ilişkisiz iki ya da daha çok örneklem ortalaması arasındaki farkın sıfırdan anlamlı bir şekilde farklı olup olmadığını test etmek için kullanılır (Büyüköztürk, 2010). Bağımsız örneklem için t testinin sahip olduğu varsayımlar tek yönü ANOVA içinde geçerlidir. Ancak bağımlı değişkene ilişkin varyansların her bir örneklem için eşit olması yani varyansların homojenliği varsayımı bağımsız örneklem için t testine kıyasla daha önemlidir ve karşılanmaması daha zor tolare edilir.

nesne adı<-aov(bağımlı değişken ~ bağımsız değişken, data=dosya adı);

summary(nesne adı): Tek yönlü varyans analizinin R ortamında hesaplanabilmesi için iki farklı fonksiyon gerekmektedir. Öncelikli olarak araştırmacı tarafından belirlenen bir nesnenin “**aov**” fonksiyonu ile oluşturulması gerekmektedir. Daha sonra oluşturulan nesne **summary()** fonksiyonu ile kullanılarak tek yönlü ANOVA sonuçları elde edilir (Crawley,2007).Örneğin “anovasonuç” nesnesinin oluşturulduğu örnek bir komut aşağıda

sunulmuştur. Komutta “örnek” isimli bir dosyada bulunan “BAŞARI” ve “SED” değişkenleri ele alınmıştır.

```
> anovasonuç<-aov(BAŞARI~SED,data=örnek)
> summary(anovasonuç)
          Df Sum Sq Mean Sq F value Pr(>F)
SED         2     53   26.482   3.199 0.0498 *
Residuals  47    389    8.277
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

Tek yönlü varyans analizi neticesinde eğer grup ortalamaları arasındaki farklar anlamlı çıkmış ise bu farkın hangi gruplar arasında olduğunun belirlenebilmesi için post hoc çoklu karşılaştırma testleri hesaplanır. R yazılımında bu işlem için farklı fonksiyonlar kullanılabilir.

pairwise.t.test(dosya adı \$ sürekli değişken adı, dosya adı \$ grup değişkeni adı, paired=F, p.adjust.method =“method”): Bu fonksiyon Bonferroni ve onunla ilişkili teknikler (Holm,Benjamini-Hochberg,Hommel vb.) kullanılarak çoklu karşılaştırma testi hesaplandığı durumlarda kullanılır. Fonksiyon içerisinde “**paired=F**” kısmı grupların bağımsız olduğu bilgisini yazılıma verir. “**p.adjust.method**” ise elde edilen p değerlerine hangi düzeltme yönteminin kullanılacağı belirtilen kısımdır. Bu kısım “**p.adjust.method=“Bonferroni”**” şeklinde yazılırsa Bonferroni yönteminin kullanılacağı bilgisi yazılıma sunulmuş olur. Bir önceki örnekte verilen değişkenler için örnek bir çoklu karşılaştırma testi için kullanılan fonksiyon ve çıktısı aşağıda sunulmuştur.

```
> pairwise.t.test(örnek$BAŞARI,örnek$SED,paired=F,p.adjust.method="bonferroni")
```

```
Pairwise comparisons using t tests with pooled SD

data:  örnek$BAŞARI and örnek$SED

      alt      orta
orta 0.206 -
üst  0.057 1.000

P value adjustment method: bonferroni
```

Örneğin fonksiyonun “**p.adjust.method**” kısmını **p.adjust.method=“BH”** şeklinde yazılırsa Benjamini-Hochberg yönteminin kullanılacağı bilgisi yazılıma sunulmuş olur. “Tukey” ve “Dunnet” gibi diğer çoklu karşılaştırma testlerinin hesaplanabilmesi için “**multcomp**” paketinin indirilmesi gerekmektedir (Field, Miles, Field, 2012).

Anlam çıkarıcı istatistikler bölümünde ele alınan analizleri sonuçlarını raporlaştırma sürecinde aritmetik ortalama, standart sapma gibi bazı betimsel istatistiklere de ihtiyaç duyulmaktadır, ancak bazı analiz çıktıları bu istatistikleri içermemektedir. Bu durumlarda ele alınan değişkenlere yönelik betimsel istatistiklerin hesaplanması için önceki bölümlerde verilen fonksiyonlardan faydalanılabilir.

Parametrik Olmayan Testler

Parametrik testlerin gerekli varsayımları karşılanmadığı durumlarda alternatif bazı parametrik olmayan testler hesaplanır. Bu bölümde “bağımsız örneklem için t testinin”, “tekrarlı ölçümler için t testinin” ve “tek yönlü varyans analizinin” hesaplanmadığı durumlarda kullanılan bazı parametrik olmayan analizlere ilişkin R fonksiyonlarına ve çıktılarına yer verilmiştir.

Wilcoxon Sıra Toplamları Testi

Bağımsız örneklem için t testinin gerekli varsayımlarının karşılanmadığı için hesaplanamadığı durumlarda kullanılan parametrik olmayan bir testtir (Siegel, Castellan,1988).Wilcoxon sıra toplamları testi Man Whitney U testi ile aynı işlem basamaklarını içermektedir ve birbirlerinin muadili niteliğindedir. Testin hesaplanabilmesi için “**wilcox.test(bağımlı değişken adı~ bağımsız değişkeni adı, data=dosya adı)**” fonksiyonu kullanılır. Aşağıda “Tutum” dosyasında bulunan öntest puanlarının cinsiyete göre anlamlı bir farklılık gösterip göstermediğine ilişkin R fonksiyonu ve çıktısı sunulmuştur.

```
> wilcox.test(ÖNTEST~CİNSİYET,data=Tutum)

Wilcoxon rank sum test with continuity correction

data: ÖNTEST by CİNSİYET
W = 14.5, p-value = 0.3252
alternative hypothesis: true location shift is not equal to 0
```

Wilcoxon İşaretli Sıralar Testi

Wilcoxon işaretli sıralar testi ilişkili iki ölçüm setine ait puanlar arasındaki farkın anlamlılığını test etmek amacıyla kullanılır (Siegel, Castellan,1988; Büyüköztürk, 2010). Gerekli varsayımları karşılanmadığı için tekrarlı ölçümler için t testinin hesaplanamadığı durumlarda parametrik olmayan alternatif bir test olarak kullanılır.Wilcoxon işaretli sıralar testi “**wilcox.test(dosya adı \$ birinci değişken adı, dosya adı \$ikincideğişken adı, paired=T)**” fonksiyonu ile R ortamında hesaplanabilir. Fonksiyonda bulunan “**paired=T**” ifadesi R yazılımına örneklemelerin ilişkili olduğu bilgisini vermektedir. Aşağıda “Tutum” dosyasında bulunan öntestvesontest puanları arasında anlamlı bir farklılık gösterip göstermediğine ilişkin R fonksiyonu ve çıktısı sunulmuştur.

```
> wilcox.test(Tutum$ÖNTEST,Tutum$SONTEST,paired=T)

Wilcoxon signed rank test with continuity correction

data: Tutum$ÖNTEST and Tutum$SONTEST
V = 0, p-value = 0.009091
alternative hypothesis: true location shift is not equal to 0
```

Kruskal Wallis Testi

Kruskal Wallis testi ilişkisiz iki ya da daha çok örneklem ortalamasının birbirinden anlamlı farklılık gösterip göstermediğinin test edilmesinde kullanılan parametrik olmayan bir istatistiktir (Büyüköztürk, Çokluk, Köklü 2008). Tek yönlü varyans analizi için gerekli varsayımlarından karşılanamadığı durumlarda sıklıkla kullanılan bir testtir. R ortamında Kruskal Wallis testinin hesaplanabilmesi için “**kruskal.test(bağımlı değişken adı~ bağımsız değişkeni adı, data=dosya adı)**” fonksiyonu kullanılır. Aşağıda sunulan fonksiyonda “örnek” isimli bir dosyada bulunan “BAŞARI” ve “SED” değişkenleri ele alınmıştır.

```
> kruskal.test(BAŞARI~SED,data=örnek)

Kruskal-Wallis rank sum test

data: BAŞARI by SED
Kruskal-Wallis chi-squared = 5.6026, df = 2, p-value = 0.06073
```

Ancak R yazılımının verdiği çıktıda Kruskal Wallis testi hesaplama sürecinde elde edilen sıra ortalamaları bulunmamaktadır. Gruplara ait sıra ortalamalarını hesaplayabilmek ve gruplar arasındaki farklılıkları görebilmek için fazladan birkaç komut yazılması gerekmektedir. Bunun için öncelikle aşağıdaki kod ile "örnek" dosyası içerisinde yeni bir değişken oluşturup sıra puanlarını hesaplamamız gerekmektedir.

```
> örnek$sıra<-rank(örnek$BAŞARI)
> örnek
      BAŞARI  SED sıra
1  5.469208  üst  14
2  4.654633  üst  12
3  6.426284  üst  20
4 11.915450  üst  45
5  5.609556  üst  17
```

"örnek" dosyasının içerisinde "sıra" isimli yeni bir değişken ve sıra puanları oluşturduktan sonra aşağıdaki fonksiyon ile her bir alt grup için sıra ortalamalarını hesaplayabiliriz.

```
> by(örnek$sıra,örnek$SED,mean)
örnek$SED: alt
[1] 32.5625
-----
örnek$SED: orta
[1] 22.88235
-----
örnek$SED: üst
[1] 21.47059
```

Anlamlı çıkan Kruskal Wallis testleri sonrasında aynı tek yönü ANOVA'da olduğu gibi farkın hangi gruplar arasında olduğunun belirlenmesi gerekmektedir. Bu durumda gruplar arasında ikişerli olarak wilcoxon sıra toplamları testinin hesaplanarak anlamlı farkın hangi grup arasında olduğu belirlenebilir. Bunun için wilcoxon sıra toplamları testi için önceki bölümlerde verilen fonksiyonlardan faydalanılabilir.

YÖNTEM

Araştırma Modeli

Türkiye'de eğitim bilimleri alanındaki ve SSCI kapsamında taranan dergilerde yayınlanan makalelerde R yazılımının ve diğer istatistiksel yazılımların kullanılma sıklığının yıllara göre belirlenmesinin amaçlandığı bu çalışma, tarama türünde bir betimsel araştırmadır. Betimsel araştırmalar verilen bir durumu olabildiğince tam ve dikkatli bir şekilde sürece müdahale etmeden olduğu gibi tanımlayan araştırmalardır. Eğitim alanındaki araştırmalarda en yaygın betimsel yöntem tarama çalışmalarıdır (Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demiral, 2009).

Çalışma Grubu

Bu araştırmada ilgili veriler; Türkiye'de eğitim bilimleri alanındaki ve SSCI kapsamında taranan 3 dergide 2010-2014yılları arasında yayınlanan 1627 makaleden elde edilmiştir. Çalışmaya TED Eğitim ve Bilim, Kuram ve Uygulamada Eğitim Bilimleri ve Hacettepe Üniversitesi Eğitim Bilimleri dergileri dahil edilmiştir. Bu dergilerin çalışmaya dahil edilmelerinin nedeni, SSCI kapsamında taranıyor olmaları ve belirli bir alandan ziyade eğitim bilimlerinin farklı alanlarına ait çalışmaları içeriyor olmalarıdır.

Aşağıda dergilere ait incelenen makale sayıları yıllara göre Tablo 1’de sunulmuştur.

Tablo 1. İncelenen dergiler ve makale sayıları

Yıllar	Türkiye’deki Dergiler			Toplam
	Eğitim ve Bilim	KUYEB	Hacettepe Ün. Eğ. Dergisi	
2010	57	63	59	179
2011	94	115	78	287
2012	96	174	130	400
2013	115	134	173	422
2014	162	100	77	339
Toplam	524	586	517	1627

Verilerin Toplanması

Çalışmada ilgili veriler çevrimiçi akademik veri tabanları ve ulaşılabilen ilgili dergilerin hardcopy’leri aracılığı ile toplanmıştır. Toplanan veriler yıllara ve dergi adlarına göre düzenlenmiş ve analize hazır hale getirilmiştir.

Verilerin Analizi

Verilerin analizi sürecinde incelenen makalelerde veri analizi için kullanılan istatistik programları; türlerine, derginin yayınlanma yılına, dergi adlarına göre yüzde ve frekanslarla belirlenmiş, yıllara yönelik değişimler grafiklerle görselleştirilmiştir.

BULGULAR

Çalışmanın bu kısmında elde edilen verilere yönelik bulgular sunulmuştur.

Türkiye’de SSCI kapsamındaki Dergilerde 2010-2014 Yıllarında Yayımlanan Araştırmalarda Kullanılan İstatistik Paket Programlarına Yönelik Bulgular

Araştırma kapsamında öncelikli olarak 2010 -2014 yılları arasında Türkiye’de SSCI kapsamında yayımlanmış 3 dergide yer alan makaleler istatistik bir paket program kullanımını gerektiren nicel çalışmalar ve istatistik kullanımı gerektirmeyen nitel veya derleme çalışmalar olarak sınıflandırılmıştır. Tablo 2 ‘de bu sınıflamaya ilişkin yıllara göre frekans ve yüzde değerleri verilmiştir.

Tablo 2: Makalelerin istatistik paket programı kullanmaya uygunluğuna yönelik frekans ve yüzdeler

Yıllar	İstatistik Paket programının kullanımı gerektirmeyen Nitel veya Derleme çalışmaları		İstatistik Paket programının kullanımını gerektiren nicel boyutu olan çalışmalar		Toplam	
	f	%	f	%	f	%
	2010	37	20.7	142	79.3	179
2011	84	29.3	203	70.7	287	100
2012	141	35.2	259	64.8	400	100
2013	135	31.9	287	68.1	422	100
2014	98	28.9	241	71.1	339	100
Toplam	495	30.4	1132	69.6	1627	100

Tablo 2 incelendiğinde araştırma kapsamında ele alınan toplam 1627 makalenin 1132 (% 69.6) tanesi istatistik paket programı kullanımı gerektiren nicel boyutu olan çalışmalardır. 495 (% 30.4) makale ise istatistik paket programı kullanımı gerektirmeyen nitel, derleme veya tartışma çalışmalarından oluşmaktadır. Yıllara göre dağılımına bakıldığında da istatistik paket programının kullanılmasını gerektiren nicel boyutlu makalelerin o yıl yayımlanan makaleler arasındaki yüzdelerinin %65 ile %79 arasında değiştiği görülmektedir. Bu bağlamda 2010-2014 yılları arasında Türkiye’de SSCI kapsamındaki 3 dergide yayınlanan makaleler arasında nicel bir boyutu olan ve istatistik paket kullanımı gerektiren makalelerin sayısının daha fazla olduğu belirtilebilir. Bulguların ileriki kısımlarında gerekli analizler istatistiksel paket programının kullanılmasını gerektiren nicel bir boyuta sahip 1132 makale üzerinden gerçekleştirilmiştir.

Analiz sürecinde kullanım sıklığı yüksek olan ve sosyal bilimler alanında genel olarak veri çözümleme amacıyla kullanılacak “SPSS”, “Lisrel”, “Amos”, “R” ve “SAS” programları dikkate alınmıştır. Kullanım sıklığı az olan ve spesifik amaçlar (Madde yanlılığı, test eşitleme, madde tepki kuramına dayalı parametre tahminleri, Genellenebilirlik kuramına dayalı parametre tahminleri vb.) için kullanılan istatistik programları (Facet, Bilog, Multilog, Genova vb.) “Diğer” başlığı altında ele alınmıştır. Bunların yanı sıra bazı çalışmalarda ise birden fazla paket programının aynı anda kullanıldığı görülmektedir. Kullanım sıklığına dayalı olarak bu tür çalışmalar SPSS ve Lisrel paket programlarının birlikte kullanımını içeriyorsa “SPSS ve Lisrel” başlığı altında SPSS ile birlikte Lisrel dışındaki başka programların kullanımını içeriyorsa “SPSS ve Diğer” başlığı altında toplanmıştır.

Tablo 3 2010 -2014 yılları arasında Türkiye’de yayımlanan SSCI kapsamındaki 3 dergide yer alan ve nicel araştırmalarda kullanılan istatistik paket programlarının türlerine ilişkin yüzde ve frekans değerleri verilmiştir.

Tablo 3: 2010 -2014 yılları arasında SSCI kapsamında giren dergilerde yayınlanan araştırmalarda kullanılan istatistik paket programlarına yönelik betimsel analizler

Program	Yıllar											
	2010		2011		2012		2013		2014		Toplam	
	f	%*	f	%*	f	%*	f	%*	f	%*	f	%
SPSS	53	37.4	66	32.5	94	36.3	117	40.8	85	35.3	415	36.7
LİSREL	-	-	5	2.4	2	0.8	4	1.4	8	3.3	19	1.7
AMOS	-	-	2	0.9	3	1	2	0.7	2	0.8	9	0.8
R	-	-	-	-	-	-	1	0.4	-	-	1	0.01
SAS	-	-	2	0.9	1	0.4	-	-	-	-	3	0.3
Diğer	3	2.1	8	3.9	5	1.9	3	1	13	5.4	32	2.8
SPSS ve LİSREL	8	5.6	10	4.9	14	5.4	12	4.2	17	7.1	61	5.4
SPSS ve Diğer	5	3.5	1	0.9	5	1.93	3	1	14	5.8	28	2.5
Belirtilmemiş**	73	51.4	109	53.6	135	52.1	145	50.5	102	42.3	564	49.8
Toplam	142	100	203	100	259	100	287	100	241	100	1132	100

*İlgili frekansın, bulunduğu sütunda yer alan yıldaki toplam frekansa göre yüzdesi

** Kullanılan istatistik paket programının belirtilmediği çalışmalar

Tablo 3 incelendiğinde ilk dikkat çekici bulgu, istatistik paket programının kullanılmasını gerektirecek nicel boyutu olan 1132 çalışmanın yaklaşık % 50’sinde, kullanılan paket programın araştırma raporunda belirtilmemiş olmasıdır. Yıllara göre bakıldığında bu oranın %42.3 ile %53.6 arasında değiştiği görülmektedir. Bu makalelerin bulgular bölümünde, paket programın kullanılmasını gerektirecek analiz sonuçları bulunmakla beraber, hangi paket programın kullanıldığı belirtilmemiştir.

İkinci dikkat çekici nokta ise 1132 makalenin yaklaşık olarak % 37’sinde SPSS paket programının kullanılmış olmasıdır. SPSS’in başka paket programlarla birlikte kullanıldığı çalışmalar da dikkate alınır, 2010-2014 yılları arasında ilgili dergilerde yayımlanmış nicel boyutu olan çalışmaların yaklaşık % 45’inde SPSS’ten yararlandığı belirtilebilir. Kullanılan istatistik programının rapor edilmediği çalışmalar dikkate alınmadığında, çalışmaların yaklaşık % 89’unda SPSS’ten faydalandığı söylenebilir.

SPSS ve Lisrel'in birlikte kullanıldığı çalışmalar %5.4 ve SPSS'in Lisrel dışındaki paket programlarla birlikte kullanıldığı çalışmalar ise % 2.5 oranına sahiptir. Başka bir ifadeyle makalelerin yaklaşık %8'inde SPSS ve başka bir istatistik paket programı birlikte kullanılmıştır. Dikkat çekici bir diğer nokta ise kullanımı giderek artan ve diğer paket programlara kıyasla belirli avantajlara sahip olan R yazılımının 1132 makaleden yalnızca 1 tanesinde kullanılmış olmasıdır. Lisrel, Amos ve SAS programları ise % 0.3 ile % 1.7 arasında değişen oranlara sahiptir. Diğer başlığı altında toplanan ve yalnızca spesifik analizler için kullanılabilen Bilog, Multilog, Genova, Facet, Ezdiff gibi programlar ise makalelerin % 2.8'ini oluşturmaktadır.

TARTIŞMA

Çalışmada elde edilen dikkat çekici önemli bulgulardan birisi dünyada oldukça sık kullanılan ve kullanım sıklığı giderek artan R yazılımının Türkiye'de 2010-2014 yılları arasında ilgili dergilerde yayınlanan makalelerde yalnızca 1 defa kullanılmış olmasıdır. Bunun yanı sıra en çok kullanılan istatistik paket programının SPSS olduğu ve 2014 yılı hariç her geçen yıl kullanım sıklığının belirlenmiştir. R yazılımının bu denli az olması ilk başta şaşırıcı gözükmeyle beraber, mevcut koşullar dikkate alındığında beklenebilecek bir durumdur. Bu durumun olası gerekçeleri şöyle sıralanabilir;

- R yazılımının ne olduğu, nasıl kullanılacağı, avantaj ve dezavantajlarının neler olduğuna ilişkin bilgi sunan Türkçe yayınlanmış bir kitap bulunmakla beraber eğitim alanında yayınlanmış hiçbir kitap bulunmamaktadır. Aynı şekilde belirtilen boyutları içeren Türkiye'de yayınlanmış sadece bir makale bulunurken eğitim bilimleri alanında hiçbir makaleye rastlanmamıştır. Bunun yanı sıra SPSS yazılımını tanıtan ve araştırmacılara kullanım yönergeleri sunan hem eğitim alanında hem de diğer disiplinlerde yayınlanmış onlarca kitap ve makale bulunmaktadır. Başka bir ifade R yazılımını kullanmak isteyen bir araştırmacının ulaşabileceği Türkçe kaynak sayısı SPSS yazılımına kıyasla oldukça sınırlıdır.
- R yazılımının diğer paket programlardan farklı olarak ilgili sekmelere tıklayarak işlemlerin yapılabilmesi bir ara yüzünün olmaması ve analizler için spesifik kod yazımını gerektirmesi, araştırmacılarda R kullanımının zor olduğuna yönelik ön yargılar oluşturmuş olabilir. Aksine SPSS yazılımı daha kullanıcı dostu bir ara yüze sahip olduğu için araştırmacılar tarafından daha çok tercih edilebilmektedir. Hâlbuki R yazılımı kısa süre kullanıldıktan sonra araştırmacılara kolay gelebilmekte ve pek çok avantaj sunabilmektedir (Field, 2009). Bu noktada R yazılımını tanıtan yayınların artmasına ve bu konuda hizmet içi eğitimler verilmesine ihtiyaç olduğu söylenebilir.
- R yazılımının tanıtımına, veri çözümleme sürecinde kullanımına yönelik bilgi ve becerileri kazandırmaya yönelik Türkiye'deki Üniversitelerin İstatistik, mühendislik, biyoistatistik gibi bölümlerinde dersler bulunmaktadır. Eğitim Enstitülerinde Lisansüstü düzeyde verilen ileri istatistik, psikometri ve madde tepki kuramı gibi derslerin içerisinde kısmen ele alınmaktadır. Ancak R yazılımının tanıtımına, nicel veri çözümleme sürecinde kullanımına ve programlamasına yönelik bir ders bulunmamaktadır.
- Bunun yanı sıra SPSS yazılımına ilişkin bilgiler araştırma yöntemleri ve istatistik dersleri kapsamında ele alınmaktadır. Bu bağlamda lisansüstü düzeyde R yazılımının kullanımına ilişkin temel bilgi ve becerileri içeren dersler açılmalı veya araştırma yöntemleri ve istatistik dersleri kapsamında elen alınmalıdır.

Üzerinde tartışılması gereken bir diğer nokta ise SPSS ve benzeri ticari yazılımların kullanımına yönelik etik sorunlardır. Bilimsel bir araştırma bulgusu bulunmamakla birlikte araştırmalarda SPSS ve benzeri yazılımların, daha ucuza temin edilen ve orijinal olmayan versiyonlarının kullanıldığı ifade edilebilir. Bu durum başlı başına etik bir sorun teşkil etmektedir. Hâlbuki ücretsiz temin edilebilen R yazılımının kullanımı, bu etik problemin aşılmasına katkı sağlayacaktır.

İncelenen makaleler arasında nicel çalışmaların yaklaşık %50'sinde veri çözümleme sürecinde hangi yazılımın kullanıldığı belirtilmemiştir. Bu durum kullanılan paket programın, veri analiz sürecinde hesaplanan istatistiksel tekniklerin önüne geçmesini engellemektedir.

Çünkü pek çok çalışmada kullanılan istatistiksel yöntemler yeterince açıklanmadan, analizlerin hangi paket program ile hesaplandığı vurgulanmaktadır. Hatta birçok çalışmada veri çözümleme yöntemleri başlığı altında ve ilk cümlede, kullanılan paket programın adı ve versiyonu sunulmaktadır. Bu bakış açısı ile kullanılan paket programının adından ziyade hangi çözümleme tekniklerinin neden ve hangi formlere dayalı olarak kullanıldığının vurgulanması önemlidir. Ancak paket program çok spesifik bir analiz için kullanılıyorsa (Facet, WinBugs, Multilog vb.) veya alanda çok yeni ve tanıtılması gereken bir yazılım ise raporlama sürecinde belirtilmesi uygun olacaktır. R ve benzeri açık kaynaklı yazılımlardan faydalanıldığı çalışmalarda, kullanılan kodların raporlanması, hem diğer araştırmacılarla paylaşılması hem de çözümleme sürecinin doğruluğunun test edilebilmesi açısından yerinde olacaktır.

KAYNAKÇA

- Baydoğan, M.G., Orbay,B. & Çetin, U. (2014). *R ile Proglamlamaya Giriş ve Uygulamalar. XIX. Türkiye’de İnternet Konferansı*. Yaşar Üniversitesi, sözlü bildiri.
- Beaujean. A. A. (2013). Factor analysis using R. *Practical Assessment: Research & Evaluation*. 18(4) <http://pareonline.net/getvn.asp?v=18&n=4>
- Büyüköztürk, Ş. Çokluk & Ö. Köklü, N. (2008). *Sosyal bilimler için istatistik* (3. baskı). Ankara: Pegem Akademi.
- Büyüköztürk, Ş., Çakmak K. E., Akgün, Ö. E., Karadeniz, Ş. & Demirel, F. (2009). *Bilimsel Araştırma Yöntemleri* (5. baskı). Ankara: PegemYayıncılık.
- Crawley, M. (2007) *The R book*. Chichester: Wiley.
- Er. F & Sönmez, H. (2005). The usage of R windows for the fundamental statistics education. *Journal of Eng.&Arch.Fac. Eskişehir Osmangazi University*, 18(2).
- Field, A, Miles. J & Field. Z. (2012). *Discovering statistics using R*. London: Sage Publications.
- Field. A. (2009). *Discovering statistics using SPSS* (2nd edition). London: Sage Publications.
- Knell, J.R (2014). *Introductory R*. United Kingdom: Hershman, Walton on Thames.
- Özdemir A.F, Yıldıztepe E. & Binar, M. (2010). İstatistiksel Yazılım Geliştirme Ortamı: R. *XII. Akademik Bilişim Konferansı* 10-12 Şubat Muğla Üniversitesi, sözlü bildiri.
- Palland, J. (2002). *SPSS survival manuel* (11th edition). PA: Open UniversityPress.
- Siegel, S. Castellan, N.J. (1988). *Nonparametric statistics for behavioral sciences* (2nd edition). New York: McGraw-Hill.
- Tabachnick, B.G. & Fidell, L.S. (1996). *Using multivariate statistics* (3rd edition). Newyork: Harper Collins.
- The Comprehensive R Archive Network. <http://cran.r-project.org>, adresinden 10 Şubat 2015 tarihinde erişilmiştir.
- Wenables, W.N. Simith, D.M. & The R Development Core Team. (2002). *An introductionto R*. Bristol: Network Theory.
- Zuur, A.F., Ieno, E.N, & Mesters, E.H.W.G. (2009). *A beginner’s guide to R*. Dordrecht: Springer-Werlag.