T.C.
İNÖNÜ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL YÖNTEMLERİN KARŞILAŞTIRMALI ANALİZİ

DOKTORA TEZİ

DANIŞMAN
Dr. Öğr. Üyesi Kamil DURDU

HAZIRLAYAN
Muhammed Bedir BAYDEMİR

MALATYA – 2020

T.C.
İNÖNÜ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL YÖNTEMLERİN KARŞILAŞTIRMALI ANALİZİ

EKONOMETRİ ANA BİLİM DALI

DOKTORA TEZİ

Muhammed Bedir BAYDEMİR

MALATYA, 2020

İNÖNÜ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL YÖNTEMLERİN KARŞILAŞTIRMALI ANALİZİ

DOKTORA TEZİ

DANIŞMAN	HAZIRLAYAN
DR.ÖĞR.ÜYESİ. KAMİL DURDU	MUHAMMED BEDİR BAYDEMİR

Jürimiz tarafından 14/02/2020 tarihinde yapılan tez savunma sınavı sonucunda bu tez oybirliği ile başarılı bulunarak Ekonometri Anabilim Dalı Doktora Tezi olarak kabul etmiştir.

Jüri Üyelerinin Unvanı Adı Soyadı	**İmza**
1. Dr. Öğr. Üyesi Kamil DURDU (Danışman)	.....
2. Prof. Dr. Hüseyin AĞIR	.....
3. Doç. Dr. Fatma ZEREN	.....
4. Dr. Öğr. Üyesi İbrahim GÖRÜCÜ	.....
5. Dr. Öğr. Üyesi Oktay KIZILKAYA	.....

O N A Y

Bu tez, İnönü Üniversitesi Lisansüstü Eğitim-Öğretim Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından kabul edilmiş ve Enstitü Yönetim Kurulu'nun .../..../20... tarih ve 20..../..... sayılı Kararıyla da uygun görülmüştür.

Prof. Dr. Mehmet KUBAT
Enstitü Müdürü

ONUR SÖZÜ

Dr. Öğr. Üyesi Kamil DURDU danışmanlığında doktora tezi olarak hazırladığım "ÇOK DEĞİŞKENLİ İSTATİSTİKSEL YÖNTEMLERİN KARŞILAŞTIRMALI ANALİZİ" başlıklı bu çalışmanın, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın tarafımdan yazıldığını ve yararlandığım bütün yapıtların hem metin içinde hem de kaynakçada yöntemine uygun biçimde gösterilenlerden oluştuğunu belirtir, bunu onurumla doğrularım.

08/01/2020

Muhammed Bedir BAYDEMİR

TEŞEKKÜR

Doktora eğitim sürecimin önemli aşamalarında ilgisi ve desteği ile her zaman yanımda olan değerli hocam, Tez Danışmanım Dr. Öğr. Üyesi Kamil DURDU'ya,

Doç.Dr. Yunus BULUT ve diğer bölüm hocalarıma,

Bana her şeyden önemli olan manevi eğitim ve terbiyeyi veren, hayatımı kazanmam için gerekli özeni gösteren anneme ve babama,

Çalışmalarımın bitmesini sabırla bekleyerek destek olan değerli eşim ve çocuklarım Meryem, Yusuf ve Ahsen'e

Teşekkürü bir borç bilirim.

ÖZET

Öğrencilerin başarı ortalaması, trafik kazası istatistikleri, tarım ve hayvancılıkta verimlilik, ekonomik veriler gibi günlük hayatta yer alan olaylarda sık sık istatistik yöntemlere gerek duyulduğu gibi önemli bilimsel çalışmaların sonuçları da istatistik yöntemlerle değerlendirilmektedir.

İstatistiksel ilişkiler en basit anlamda bir bağımsız değişkenin bir bağımlı değişkeni etkilemesi temeline dayanmaktadır. Oysa doğa olaylarından sağlık alanındaki olaylara kadar değişken veya değişkenleri etkileyen birden çok faktör vardır. Tek değişkenli istatistiksel analizlerin eksikliği ve sınırlı olayları açıklayabilmesi, araştırmalarda çok değişkenli istatistiksel analiz yöntemlerin kullanılmasını gerektirmiştir.

Çok değişkenli istatistik yöntemlere alternatif olarak tek değişkenli yöntemlerin art arda uygulanması da düşünülebilir. Ancak bu durum birçok yöntem için değişkenler arasındaki etkileşimin ihmal edilmesine ve aynı zamanda tesadüfi hata oranlarının da artmasına neden olacaktır. Tek değişkenli hipotez testlerin art arda uygulanması ile çok değişkenli hipotez testlerinin sonuçları da aynı olmayabilir. Örneğin normallik testi için değişkenler tek tek test edildiğinde sıfır hipotezi kabul edilerek tüm değişkenlerin normal dağılıma uyduğu kabul edilebilir. Ancak değişkenler birlikte çok değişkenli normalliği sağlamayabilirler.

Çok değişkenli istatistik yöntemler, bilginin birikimli olarak ilerlemesi ilkesiyle ihtiyaçları karşılamak üzere geliştirilmişlerdir. Tamamen aynı amaçla kullanılacak, tüm varsayımları da aynı olacak yeni bir yönteme ihtiyaç olmayacağı aşikardır. Her yöntem, varsayım ve göstergelerine göre özgün olacağından çok değişkenli istatistikler amaç ve yöntemlerine göre kesin bir sınıflandırmaya ayrılamazlar.

Bu tez çalışmasında çok değişekenli istatistik ve bunların varsayımlarının sınanmasında kullanılan bazı yöntemlerin karşılaştırması yapılmıştır. Değişkenlerin ayrı ayrı tek değişkenli normalliği sağlamalarının çok değişkenli normalliği sağlamak için ölçü olup olmayacağı test edilmiştir.

ANAHTAR KELİMELER: Çok Değişkenli İstatistik Yöntemler, Çoklu Normal Dağılım, Diskriminant Analizi, Lojistik regresyon Analizi, Probit Analizi, Kümeleme Analizi, Faktör Analizi

ABSTRACT

Statistical methods; the results of important scientific studies are also evaluated by statistical methods, such as the average achievement of students, traffic accident statistics, productivity in agriculture and animal husbandry, economic data, as well as frequently needed events in daily life.

In the simplest sense, statistical relations are based on the effect of an independent variable on a dependent variable. However, there are multiple factors that affect the variables from natural events to health events. The lack of univariate statistical analyzes and its ability to explain limited events required the use of multivariate statistical analysis methods.

As an alternative to multivariate statistical methods, the application of univariate methods consecutively can be considered. However, this will neglect the interaction between variables for many methods, and also lead to an increase in random error rates. The results of univariate hypothesis testing may not be the same with successive multivariate hypothesis testing. For example, if the variables are tested individually for the normality test, the null hypothesis is accepted and all variables can be considered to conform to the normal distribution. However, variables may not provide multivariate normality together.

Multivariate statistical methods have been developed to meet the needs with the principle of cumulative progression of information. Obviously, there will be no need for a new method that will be used for the same purpose and all assumptions will be the same. Since each method will be unique according to its assumptions and indicators, multivariate statistics cannot be separated into a precise classification according to their aims and methods.

In this thesis, a comparison of some of the methods used to test multivariate statistics and their assumptions has been made. It has been tested whether the variables provide individual univariate normality or not to provide multivariate normality.

KEYWORDS: Multivariate Statistical Methods, Multiple Normal Distribution, Discriminant Analysis, Logistic Regression Analysis, Probit Analysis, Cluster Analysis, Factor Analysis

İÇİNDEKİLER

TABLO DİZİNİ

ŞEKİL DİZİNİ

1. GİRİŞ

İstatistik yöntemlere öğrencilerin başarı ortalaması, trafik kazası istatistikleri, tarım ve hayvancılıkta verimlilik, ekonomik veriler gibi günlük hayatta yer alan olaylarda sık sık gerek duyulduğu gibi önemli bilimsel çalışmaların sonuçları da istatistik yöntemlerle değerlendirilmektedir. Tek kalem ürün pazarlayan küçük bir işletme sahibi için günlük ortalama satış miktarını belirlemek temel bir istatistik yöntem olduğu gibi sosyal bir konuda hazırlanan anket sonuçlarının değerlendirmesi de istatistiksel yöntemlerle yapılmaktadır. Bunun yanında tıp alanında yapılan bir çalışmanın sonuçlarının anlamlılığı da yine istatistik yöntemlerle test edilmekte ve hatta hastalık teşhislerinde istatistik yöntemlerin önemli bir yeri vardır. Birçok hastalık teşhisinde biyopsi kullanılmaktadır. Ancak biyopsi invaziv bir yöntem olup cerrahi müdahale gerektirmektedir. Vücut sıvılarındaki bazı değerler veya organların büyüklük ölçülerinin istatistiği ile hastalık teşhisinin maliyeti daha düşük ve ileri uzmanlık bilgisi gerektirmemektedir. Bu şekilde tanı konulması tıpta sürekli tercih edilen güncel konularından biridir (Alpar, 2017). Kısaca istatistiğin meteoroloji, tarım, hayvancılık, her türlü ticari faaliyet, sağlık, sosyal alanlar, psikoloji vb. kullanılmadığı alan hemen hemen yok gibidir.

Çok geniş bir kullanım alanına sahip olmasına rağmen istatistiğin bilim dalı olup olmadığı hakkında farklı görüşler mevcuttur. Bilim dalı olmadığını belirtenler, istatistik yöntemlerin çok farklı alanlarda kullanılan, birçok bilim dalında ihtiyaç duyulan analiz yöntemleri olup kendine özgü ve bütünlük arz eden bir konusu bulunmadığını belirterek bilim dalı olmadığını savunmaktadırlar. Bilim dalı olduğu düşüncesinde olanlar ise istatistik yöntemlerin yerinin başka türlü doldurulamadığını, istatistik yöntemlerin kullanılmadığı durumlarda diğer birçok araştırma sonuçlarının yorumlanamayacağı şeklinde görüş belirtmektedirler. Sonuç olarak bilim olarak kabul edilmese dahi hem günlük hem de bilimsel araştırmalarda istatistik yöntemlere her zaman her alanda az veya çok ihtiyaç duyulmaktadır (Akdeniz, 2015; Tuaranlı ve Güriş, 2015).

Bilimsel bir çalışma yapılırken öncelikle problem tespit edilir. Probleme ilişkin gözlemler yapılır, veriler toplanır, veriler işlenerek bilgi haline getirilir. Buraya kadar yapılan işlemler ne kadar özenli olursa olsun istatistiksel sonuçları doğru yorumlanmayan bir araştırma hiçbir probleme çözüm olamaz. Bu eksiklik amaca ve

veriye uygun istatistiksel yöntemin seçilmesiyle giderilebilir. Uygun istatistiksel testin seçimi ise araştırmanın amacına, değişken türlerine, kullanılan veri tiplerine, gözlem sayısına ve gerekli varsayımların sağlanıp sağlanmaması gibi birçok etkene bağlıdır. Bu nedenle kullanılacak istatistiksel analiz yönteminin belirlenmesi ve araştırma sonunda elde edilen analiz sonuçlarının yorumlanması oldukça önem arz etmektedir. Teknolojinin gelişimiyle birlikte günümüzde istatistik paket programları oldukça yaygınlaşmış olsa da yapılan çalışmanın doğru istatistiksel sonuçları için araştırmacının istatistik bilgisi de oldukça önemlidir. Bu bağlamda araştırmalarda en önemli hususlardan biri de araştırmacının bilgi birikimi, tecrübesi, konuya olan hakimiyet düzeyidir. Sonucun doğru olması için öncelikle araştırmacı neyi, niçin araştırdığını iyi bilmeli, değişkenlerini ona göre seçmeli, çalışmasına uygulayabileceği istatistik yöntemleri önceden düşünmeli, hangi aşamada hangi sorunlarla karşılaşabileceğini ve bunların üstesinden nasıl gelebileceğini bilmelidir.

Araştırmalarda, eldeki verilere hangi istatistik yöntemlerin uygulanabileceğini belirlemek için bazı kriterler vardır. Doğru yöntemlerle araştırmayı ilerletmek, araştırmanın güvenirliğini artırarak sonuçların da tutarlı olarak yorumlanmasını sağlamaktadır. Araştırma için veriler elde edildikten sonra en önemli soru "elimizdeki veri için en uygun yöntem hangisi" sorusudur. Bu sorunun cevaplanması için yine cevaplanması gerekli olan sorular vardır. Kullanılacak yöntem; değişkenlerin ölçek türüne, değişken sayısına, grup sayısına hatta araştırılan konunun kendisine göre farklılık göstermektedir (Türkçe İstatistik Rehberi, 24.10.2018-web).

Parametrik yöntem uygulanması düşünülen bir yöntem için veriler toplandıktan sonra verilerin bazı varsayımları sağlamadığı görülebilir. Bu gibi durumlar için alternatif yöntemler veya dönüşümler ya da çözümler önceden planlanmalıdır. Nihayetinde istatistik matematiğe dayalı yöntemlerdir. Paket programlarına girilen herhangi bir veri topluluğu için yanlış da olsa bir sonuç çıkacaktır. Bu sonuca göre yapılan yorumlar da elbette yanlış olacaktır. Yanlışlık fark edilse dahi çalışmaya uygulanacak istatistik yöntem bulunmaması halinde yapılan ön çalışmalar, verilerin toplanma-işlenme süreçleri, maliyet gerektiren bir çalışma ise maddi kayıplar ve en önemlisi boşa geçirilmiş zaman olacaktır. İstatistik analizlerin kullanımının üzerinden yüz yıllar geçmiş olmasına rağmen günümüzde dahi yapılan birçok çalışmada doğru istatistiksel yöntemin kullanılmadığını gösteren araştırmalar vardır. Yeterince

bilinmeyen istatistik kavramlar ve bu nedenle kullanılan yanlış bilimsel yöntemler nedeniyle literatürün yarısına yakınında en az bir istatistiksel hatanın olduğu belirtilmektedir. Bu da bilim dünyası için oldukça tehlikeli olduğundan bu tür hataların azaltılması hatta tamamen kaldırılması adına birçok çalışma yapılmıştır (Sayın, 2010; Karaağaoğlu, 2005; Yücel Toy ve Güneri Tosunoğlu, 2007; Sayın, 2012).

Doğru teste karar verebilmede ilk olarak istatistiksel kavramların tam olarak anlaşılması gerekmektedir. Ana kütle, örneklem, veri, birim vb. temel kavramlardan parametrik, nonparametrik, normallik, homojenlik gibi daha ileri kavramlara kadar tam olarak özümsenmelidir. Örneğin kullanılan en basit veri türü olan kategorik verilerin matematiksel eşitlikle kodlanması sembol olup harflerle de kodlama yapılabilir veya istatistiğin doğası gereği istatistik yöntemlerle elde edilen fonksiyonlardaki eşitlik matematikte kullanılan eşitlik ile aynı değildir. Matematikte gerçek değerine çok yakın olan bir değer dahi tam olarak eşit kabul edilemez. Sembol olarak da "eşittir"den farklı olarak "yaklaşık olarak eşittir" kullanılması gerektiği bilgisine sahip olmayan araştırmacı problemin başından itibaren birbirini takip eden yanlış yorumlara neden olacaktır. Hatta istatistik yöntemler sonucunda elde edilen fonksiyonlar her zaman geçerli, genel kurallar da değildir. Sosyal, ekonomik, teknolojik gelişmelere göre yere, zamana, araştırmacının ve gözlemlerin eğilimlerine göre birçok faktörden etkilenerek değişebilir (Gürsakal, 2015).

Bu çalışmada çok değişkenli istatistik yöntemlerden benzer olanlardan bazılarının karşılaştırması yapılmıştır. Parametrik yöntemler için gereken, normal dağılıma uygunluk kriterinden dolayı öncelikle normal dağılım varsayımına ilişkin karşılaştırmalar yapılmıştır. Çok değişkenli normallik, tek değişkenli normallik ile ilişkili olduğundan teker teker normal dağılıma uygun olan değişkenler için Mardia'nın çok değişkenli normallik testi hesaplanmış ve yorumlanmıştır. Çok değişkenli normalliği sağlayan veri kümesindeki değişkenlerin ayrı ayrı tek değişkenli normalliği de sağlayacağı görüşüne ilişkin karşılaştırmalar yapıldığı gibi bunun tersine ilişkin karşılaştırmalar da yapılmıştır. Bu karşılaştırma yapılırken tek değişkenli normalliği inceleyen yöntemler arasında da karşılaştırmalar yapılmıştır. Normallik için en çok kullanılan yöntemlerden bazılarının değişkenleri normal dağılıma uygun kabul etme düzeyleri kıyaslanmıştır. Çok değişkenli istatistik yöntemlerinin temeli, değişkenler arasındaki ilişkilere dayalı olup, en basit anlamda korelasyon katsayısıları ile

incelenebilir. Normallik varsayımı ve veri türlerine göre kullanılan korelasyon katsayısı karşılaştırmaları da yapılmıştır.

Bağımlı değişkenin nitel veri olması durumunda kullanılan diskriminant analizi, lojistik regresyon analizi ve probit analizi yöntemlerinin modele aldıkları değişkenlere verdikleri önem derecelerine göre karşılaştırmaları yapılmıştır. Diskriminant analizi ve lojistik regresyon analizlerinin sınıflandırma oranlarındaki başarıları karşılaştırılmıştır. Bu yöntemler karşılaştırılırken aynı yöntemlere ait farklı teknikler olan tam ve adımsal metotlara ilişkin karşılaştırma da ayrı bir başlık olmaksızın verilmiştir. Bu üç yöntemi bir arada karşılaştıran çalışmaların olmadığı gözlemlenmiş olup, ikişer ikişer karşılaştırma yapan çalışmaların da sayısının oldukça az olduğu gözlemlenmiştir.

Faktör analizi ve kümeleme analizi gruplandırma yapmaları açısından benzer yöntemler olarak görünmektedirler. Daha çok değişkenleri gruplandırmada kullanılan faktör analizi ile gözlemleri gruplandırmada kullanılan kümeleme analizi yöntemlerinin karşılaştırması yapılmıştır. Karşılaştırmaları yapılan yöntemlerin kaynaklarda yer aldığı ancak aynı veriler üzerinde karşılaştırmalı olarak yer almadığı gözlemlenmiştir.

Bu çalışmanın içeriğine alınmasa da birçok veri kümesi üzerinde çalışılmış, farklı durumların sonuçları kıyaslanmıştır. İstenilen ön şartları taşıması gereken verilerde simülasyon yöntemlerinden de yararlanılmıştır. Ancak sonuç ve yorum olarak gerçekle uyumludurlar. Analizler SPSS 21 programı ile yapılarak, elde edilen sonuç ve tablolar paylaşılmıştır. Çok değişkenli normallik varsayımı için Mardia'nın çok değişkenli normallik testi, web tabanlı program kullanılarak belirlenmiştir.

Bu çalışmanın giriş olan birinci bölümünde istatistiğin tanımı, gerekliliği, kullanım alanları, diğer bilim dallarındaki yeri ve önemi kısa bir özet olarak verilmiştir. İstatistik yöntemlerin temeli olan tek değişkenli istatistik yöntemlerden çok değişkenli istatistik yöntemlere geçişin zorunluluğu, amacı ve yararlarından bahsedilmiştir. İkinci bölümde çok değişkenli istatistik yöntemler, biraz daha ayrıntıya girilerek anlatılmış, tanımı, işleyiş süreçleri, varsayımlarına, değinilmiştir. Üçüncü bölümde bu varsayımlardan normal dağılıma ilişkin karşılaştırmalar yapılmıştır. Tek değişkenli normal dağılım testleri arasında karşılaştırmalar yapıldığı gibi, veri kümesindeki değişkenlerin tek değişkenli normal dağılımı sağlamalarının çok değişkenli normal dağılım için gerek ve yeter şart olup olmama durumları da incelenmiştir. Dördüncü bölümde korelasyon katsayısı karşılaştırmaları yapılarak sonuçları paylaşılmıştır.

Beşinci bölümde nitel bağımlı değişkene sahip veri analizlerinde kullanılan diskriminant analizi, lojistik regresyon analizi ve probit analizi farklı varsayımları sağlayan veriler için karşılaştırılmıştır. Altıncı bölümde boyut indirgeme yönleriyle birbirlerine benzeyen faktör ve kümeleme analizleri karşılaştırılmıştır. Yedinci bölümde ise bu çalışma ile elde edilen sonuçlar, gözlemler, değerlendirmeler ve öneriler paylaşılmıştır.

2. ÇOK DEĞİŞKENLİ İSTATİSTİK YÖNTEMLER

İstatistik ilişkiler en basit anlamda bir bağımsız değişkenin bir bağımlı değişkeni etkilemesi temeline dayanmaktadır. Oysa doğa olaylarından sağlık alanındaki olaylara kadar değişken veya değişkenleri etkileyen birden çok faktör vardır. Bu durum araştırma yapılırken olayların açıklanmasında tek değişkenli istatistiklerin yetersiz ve eksik kalmasına neden olmaktadır. Tek değişkenli istatistiksel analizlerin eksikliği ve sınırlı olayları açıklayabilmesi, çok değişkenli istatistiksel analiz yöntemlerin kullanılması gereksinimine neden olmuştur. Araştırmalarda daha tutarlı sonuçlar elde etmek için çok değişkenli istatistiksel yöntemlerin kuramsal gelişimi günümüzde de devam etmektedir (Albayrak, 2003).

Çalışmalarda genellikle incelenen değişkenlerin birçoğu birbirinden bağımsız değildir, bir veya daha fazla değişken ile ilişki içindedirler. Bundan dolayı herhangi bir değişken incelenirken, bu değişken ile ilişkili diğer tüm değişkenleri ya sabit kabul etmek ya da kontrol altına almak gerekmektedir. Bu da doğru çözüme ulaşmaya engel olmaktadır. Çok değişkenli istatistik yaklaşımlar, çok sayıdaki değişkenin oluşturduğu veri yapısını basit bir forma dönüştürüp problemin yapısına uygun çözümler için daha sade bilgiler ortaya koymaktadır (Sağlam, 2013).

Çok değişkenli istatistik yöntemlere alternatif olarak tek değişkenli yöntemlerin art arda uygulanması da düşünülebilir. Ancak bu durum birçok yöntem için değişkenler arasındaki etkileşimin ihmal edilmesine ve aynı zamanda tesadüfi hata oranlarının da artmasına neden olacaktır. Ayrıca tek değişkenli hipotez testlerin art arda uygulanması ile elde edilen sonuçlar ile çok değişkenli hipotez testlerinin sonuçları arasında benzerlik de bulunmayabilir. Örneğin normallik testi için değişkenler tek tek test edildiğinde sıfır hipotezi kabul edilerek tüm değişkenlerin normal dağılıma uyduğu kabul edilebilir. Ancak değişkenler hep birlikte çok değişkenli normalliği sağlamayabilirler (Kalaycı, 2010; Albayrak, 2003).

Çok değişkenli istatistik için ortak bir tanım bulunmamakta olup, yazar ve araştırmacılar tarafından çeşitli tanımlar yapılmıştır. Shin'e göre çok değişkenli istatistiksel analiz, eş zamanlı olarak çok sayıda bağımlı değişken ya da bağımlı/bağımsız değişken ayrımı yapılmaksızın çok sayıda değişkenle ilgilenildiğinde başvurulan yöntemlerdir. Sheth'e göre çok değişkenli analiz, örnek üzerinde ikiden

fazla değişkeni eş zamanlı çözümleyen tüm istatistik tekniklerdir. Gatty "değişken grupları arasındaki karşılıklı ilişkileri ölçme ve açıklama imkanı veren tüm istatistik teknikler" olarak tanımlamıştır. Hair ve arkadaşları ise çok değişkenli analiz; çoklu değişkenlerin tek bir ilişki veya ilişki kümesi içerisindeki analizidir tanımını yapmışlardır. Timm'e göre, bağımlı ve bağımsız değişken kümeleri arasında ilişki kurmak amacıyla kullanılan analizlerdir. Harris çok değişkenli istatistik için değişken kümelenmesinin bir tahmin edici veya performans ölçütü olarak dahil edildiği durumları ele almak amacıyla geliştirilmiş, açıklayıcı ve anlaşılan tekniklerin bir sınıflandırmasıdır demiştir. Afifi ve Clark çok değişkenli analizi, çalışma konusu olan her bir kişi veya birim için çok sayıda değişkenin elde edildiği verilerin analizlerini açıklamak amacıyla kullanılmasıdır diye tanımlamıştır. Kachigan'a göre çok değişkenli istatistiksel analiz, bir nesne kümesi ile ölçülen iki veya daha fazla değişken özelliklerinin aynı anda araştırılması için çalışan istatistiksel analiz dalıdır. Shaw, "iki veya daha fazla değişkeni aynı anda analiz etmeye yarayan yöntemleri tanımlamakta kullanılan genel bir terim" olarak tanımlamıştır (Albayrak, 2003; Ünlükaplan, 2008).

Çok değişkenli istatistiksel analizde sistem içinde birbiri ile ilişkili çok sayıda değişken söz konusudur. Bu analizlerde amaç, kullanılacak tekniklerle söz konusu sistemin yapısının belirlenmesi ve basit bir forma dönüştürülmesidir (Tatlıdil, 2002). Çok değişkenli istatistiksel analizler, araştırılan konuyla ilgili çok sayıda iç ve dış faktörleri göz önüne alarak, problemin yapısına ilişkin bilgiler çerçevesinde incelemek ve çözüme ulaşmak için geliştirilmiştir (Özdamar, 1999). Çok değişkenli istatistik analizler, birçok değişken arasındaki karmaşık ilişkilerin yorumlanmasına imkân sağlamaktadır (Erçetin, 1993). Çok değişkenli istatistik teknikler değişkenler arasındaki karmaşık ilişkilere açıklık getirirler (Biçkici, 2007). Yukarıdaki tanımlar da göz önüne alındığında tek değişkenli istatistiksel analizlerde veri olarak kabul edilen birçok faktörün, çok değişkenli analizlerde birer değişken olarak değerlendirilerek çalışmalar yapıldığı da söylenebilir (Ünlükaplan, 2008).

Çok değişkenli istatistik yöntemler, avantajları nedeniyle oldukça geniş bir kullanım sahasına sahiptirler. Albayrak (2005), Türkiye'de illerin sosyo-ekonomik gelişmişlik düzeylerini çok değişkenli istatistik yöntemlerle incelemiştir. Ünlükaplan (2008), peyzaj ekolojisi araştırmalarında çok değişkenli istatistiksel yöntemleri kullanmıştır. Arslan (2008), çok değişkenli istatistik analiz ile su kalitesi üzerinde

çalışmıştır. Çiftçi (2008), kalkınma göstergelerinden olan ortalama yaşam beklentisini çok değişkenli istatistik yöntemleri kullanarak Türkiye ile Avrupa Birliğini karşılaştıran çalışma yapmıştır. Bayata ve Hattatoğlu (2010), yapay sinir ağları ve çok değişkenli istatistik yöntemler ile trafik kazaları üzerine modelleme çalışmışlardır. Öztürk ve Türker (2010), Devlet Orman İşletmeleri'ni gruplandırmada çok değişkenli istatistik analizlerden yararlanmışlardır. Can (2011), bazı çok değişkenli istatistiksel teknikler arasındaki ilişkileri araştırmış ve uygulamalar yapmıştır. Sağlam (2013), toprak özelliklerini gruplandırmak için çok değişkenli istatistiksel yöntemleri kullanmıştır. Uysal ve Ersöz (2017), Türkiye'de illerin yaşam endeksini çok değişkenli istatistik yöntemleri kullanarak incelemişlerdir. Toktay (2017), çok değişkenli istatistik analiz yöntemleri olan faktör analizi ve diskriminant analizi ile üniversite öğrencileri üzerine uygulama çalışması yapmıştır. Akdamar (2018), çok değişkenli istatistik teknikler ile akıllı kentlere ilişkin çalışma yapmıştır. Kanonik korelasyon analizi ve veri zarflama yöntemleri için uygulama yapmıştır.

2.1. Çok Değişkenli İstatistik Yöntemlerde Verilerin Gösterimi

Çok değişkenli istatistiksel analizlere ait veriler tek değişkenli analiz yöntemlerinden farklı olarak genellikle n tane gözlem ve p tane değişkenden oluşmaktadır. Bu gösterim Tablo 2.1'deki gibi nxp boyutlu matris oluşturur.

Tablo 2.1. Çok Değişkenli İstatistiklerde Verilerin Gösterimi

	Değişkenler				
Gözlemler	1. değişken	2. değişken	…	…	p. değişken
1. gözlem					
2. gözlem					
…					
…					
n. gözlem					

Tek değişkenli istatistik yöntemler aritmetik ortalama, standart sapma, varyans gibi tek bir ölçü ile tanımlanmakta ve veri yapısı incelemektedir. Çok değişkenli istatistik yöntemlerde ise benzer olarak ortalama vektörü, kovaryans, varyans-kovaryans

matrisi, korelasyon katsayısı ve korelasyon matrisinden faydalanarak değişkenlikler hakkında yorum yapılmaktadır.

Tek değişkenli analizlerde verilerin toplamının veri sayısına bölünmesi ile tek bir ortalama bulunuyorken, çok değişkenli yöntemlerde bu işlem her bir değişken için yapılarak p tane değişkenin her biri için ortalama bulunmaktadır. Ortalama vektörü denilen bu değerler px1 tipinde matris oluşturmaktadır. Bazı çalışmalarda ise gözlemlere ilişkin ortalama hesabı gerekebilir. Bu durumda her bir gözlemin her değişken için aldığı değerlerin ortalaması hesaplanır.

Tek değişkenli analizlerde standart sapma ve varyans, ortalamadan sapmanın yani dağılımın yaygınlığı için birer ölçüttür. Tek değişkenli yöntemlerde bir tane varyans değeri bulunurken çok değişkenli yöntemlerde her bir değişken için benzer hesaplamalar yapıldığından p tane değişken varsa p tane varyans değeri olmaktadır. Bu değerler pxp tipinde yani değişken sayısı kadar karesel bir matris oluşturur. Her bir değişkenin diğer tüm değişkenlerle olan kovaryansı hesaplanırken değişkenin kendisiyle olan kovaryansı, varyansına eşit olacağından bu matris ile esas köşegeni her değişkenin varyansına eşit olan pxp tipinde simetrik bir matris elde edilir. Köşegenini varyansların, diğer kısımlarını simetrik kovaryansların oluşturduğu Tablo 2.2 ile gösterilen bu matrise varyans-kovaryans matrisi denir.

Tablo 2.2. Varyans-Kovaryans Matrisi Gösterimi.

	$\mathbf{X_1}$	$\mathbf{X_2}$	**…**	$\mathbf{X_p}$
$\mathbf{X_1}$	$Var(X_1)$	$Kov(X_1X_2)$	…	$Kov(X_1X_p)$
$\mathbf{X_2}$	$Kov(X_2X_1)$	$Var(X_2)$	…	…
…	…	…	…	…
$\mathbf{X_p}$	$Kov(X_pX_1)$	…	…	$Var(X_p)$

2.2. Çok Değişkenli İstatistiksel Yöntemlerin Uygulanma Amaçları ve Süreçleri

Çok değişkenli istatistik yöntemler, bilginin birikimli olarak ilerlemesi ilkesiyle ihtiyaçları karşılamak üzere geliştirilmişlerdir. Tamamen aynı amaçla kullanılacak, tüm varsayımları da aynı olacak yeni bir yönteme ihtiyaç olmayacağı aşikardır. Her yöntem, varsayım ve göstergelerine göre özgün olacağından çok değişkenli istatistikler amaç ve yöntemlerine göre kesin bir sınıflandırmaya ayrılamazlar. Bu durum istatistiğin kesin

ilişkiler değil, stokastik ilişkiler ile ilgilenmesi gerçeğiyle de bağdaşmaktadır. Örneğin bağımlı değişken tek ve sürekli ise çoklu doğrusal regresyon analizi akla gelirken, bağımlı değişken kategorik ise doğrusal olasılık modeli, probit model veya lojistik regresyon analizi kullanılabilir. Bağımlı değişken kategorik olduğunda ayırma analizi de düşünülebilir. Ancak bu durumda ayırma analizi varsayımlarını test etmek gerektiği unutulmamalıdır. Bağımlı değişken birden fazla ise bu defa kanonik korelasyon analizi düşünülmelidir. Varsayımların sağlanması durumunda parametrik yöntemlerin parametrik olmayan yöntemlerden daha iyi sonuç vereceği savına göre varsayımların sağlanması halinde diskriminant analizi, lojistik regresyon analizi ve probit analizinden daha iyi sonuç vermelidir. Kısaca her analiz türünün farklı bir tercih nedeni ve sonuçları vardır. Bunun sonucu olarak bir araştırmada kullanılabilecek birden fazla, çok değişkenli istatistik yöntem de olabilir. İlk olarak araştırmanın önceliğine göre kullanılacak yönteme karar verilir.

İstatistik yöntemler genel olarak aşağıdaki amaçlarla kullanılmaktadırlar:

a. Boyut İndirgeme: Elde edilen verilerin yapısını aralarında ilişki bulunmayan daha az sayıda değişkenle açıklamak.

b. Kümelendirme ve Sınıflandırma: Veri seti içerisinde birbirine benzer gözlemleri aynı kümelerde toplamak.

c. Bağımlılık Yapısının İncelenmesi: Değişkenler arasındaki ilişkilerden yararlanarak bağımlılığı araştırmak.

d. Atama ve Ölçekleme: Birimlerin belli ölçülere göre atanmasının yanı sıra çok sayıda değişkenden faydalanılarak birimlerin daha küçük boyutlarla gösterilmesini sağlamak.

e. Hipotez Testleri ve Hipotez Oluşturma: Tek değişkenli istatistiksel yöntemlere benzer şekilde, kurulacak hipotezleri test etmek için kullanmak (Tatlıdil, 2002).

Değişkenler arasında ilişki olması halinde, çok değişkenli istatistiksel modeller bağımlı ve iç bağımlı modeller olarak sınıflandırılabilir. Bağımlı modeller ile analiz sonucunda bağımsız değişkenlerden bağımlı değişkene ilişkin tahminler yürütülür. İç bağımlı modellerde ise bağımlı/bağımsız değişken ayrımı yapılmadan iç ilişkiler belirlenmeye çalışılır. Çok değişkenli bağımlı yöntemlere lojistik regresyon analizi, kanonik korelasyon analizi, ayırma analizi, çoklu regresyon analizi, doğrusal olasılık

modelleri ve MANOVA, iç bağımlı yöntemlereyse faktör analizi, uyum analizi ve kümeleme analizi örnek verilebilir (Ünlükaplan, 2008).

Çok değişkenli istatistiklerde amaca yönelik hangi analizlerin yapılacağının genel olarak Tablo 2.3'deki gibi olduğu söylenebilir.

Tablo 2.3. Analizlerin Amaçları

No.	Amaç	Yöntem
1	Tahminde bulunmak	Çoklu regresyon analizi, ayırma analizi, lojistik regresyon analizi
2	Sonuç çıkarmak	Çoklu regresyon analizi, MANOVA, ayırma analizi, faktör analizi, lojistik regresyon analizi
3	Sınıflama yapmak	Ayırma analizi, lojistik regresyon analizi, kümeleme analizi
4	Atama belirlemek	Ayırma analizi, temel bileşenler analizi, faktör analizi, kanonik korelasyon analizi, çok boyutlu ölçekleme analizi, uyum analizi

Çok değişkenli istatistik yöntemlerin en sık kullanım nedenlerinden biri de sınıflandırma yapma ihtiyacıdır. Gözlem yapılan yeni bir birimin hangi sınıfa dahil olacağını en az hata ile belirlemek araştırmacı için oldukça önemlidir. Belli sayıdaki özellik yani değişken incelenerek yeni gelen birimin grubu belirlenebilir. Sınıflandırma, doktorun yeni gelen bir hastanın laboratuvar sonuçlarını inceleyerek hasta olup olmadığına karar vermesi gibi, bir gözlemi uygun koşullar altında ait olabileceği en uygun sınıfa atama işlemidir.

İstatistikte bazı durumlar için olasılık dağılımları ve bunlara ilişkin parametreler bilinmektedir. Ancak bilinen bu dağılımlar, artık çok geniş bir uygulama alanı olan istatistik yöntemler için her zaman kullanılabilir olmayabilirler. Çalışmaların çoğunluğunda veriler, standart bir dağılıma tam olarak uymayan, araştırmaya özgü materyal ve değişkenlerdir. Bu durumda parametreler, alınan örneklerden elde edilen sonuçlardan çıkartılmaya çalışılır. Bu çalışmalar ile grupların ayırt edici özellikleri belirlenebilmekte ve iyi bir sınıflandırma yapılabilmektedir. Hatta analiz sonucu elde edilen sayısal değerler ile değişkenlerin birbirlerine olumlu veya olumsuz yönde etkileri de yorumlanabilmektedir. Bunun yanında elde edilen fonksiyonlar yardımı ile gözlemleri sınıflandırma imkanı da elde edilmektedir. Sınıflandırma iki şekilde olabilir. Önceden belirlenmiş gruplara atama yapılabileceği gibi çalışmanın esas amacının

grupları belirlemek olan sınıflandırma da olabilir. Grupların önceden bilinmesi durumunda diskriminant analizi ve lojistik regresyon analizi kullanılabilir. Kümeleme analizi ve çok boyutlu ölçekleme analizi ise grupları belirleme ve sınıflandırma yöntemlerine örnek verilebilirler (Burmaoğlu vd., 2009).

3. ÇOK DEĞİŞKENLİ İSTATİSTİK YÖNTEMLERDE VARSAYIMLAR

Çok değişkenli istatistiksel yöntemlerin uygulanabilmesi için bazı ön koşullar gerektirmeyenleri olduğu gibi bazı varsayımları sağlaması gereken yöntemler de vardır. Varsayımları sağlaması gereken yöntemlerin, varsayımları sağlayıp sağlamadığı analiz öncesi incelenmelidir (Albayrak, 2003).

3.1. Çoklu Normal Dağılım ve Mardia'nin Çok Değişkenli Normallik Testi

Parametrik yöntemlerin varsayımlarından en önemlisi, verilerin normal dağılıma uygun olmasıdır. Tek değişkenli normal dağılım için Kolmogorov-Simirnov, Lilliefors, Shapiro- Wilk, Çarpıklık-Basıklık, Ki-Kare Uygunluk gibi analitik testler yanında görsel olarak da karar verilebilecek histogram, dağılış poligonu, kutu-çizgi grafiği, dal-yaprak, Q-Q, P-P, detrended gibi grafik yöntemler kullanılabilir.

Çok değişkenli istatistiksel yöntemler için de en önemli varsayımlarından biri verilerin çoklu normal dağılıma uymasıdır. Temel bileşenler analizi, ayırma analizi, MANOVA, kanonik korelasyon gibi yöntemlerde bu varsayımın sağlanması gerekir. Varsayım sağlanmadığında ise yöntemlerin performansları ve güçleri azalır. Bu nedenle, varsayımın kontrolü, sonuçların anlamlı olması için önemlidir.

Literatürde 50'den fazla yöntem bulunmaktadır. Ancak, genel geçer standart bir test yoktur. Farklı yöntemlerle farklı sonuçlar elde edilebilir. Sonuçların daha güvenilir olmasını sağlamak için birden fazla test uygulanması ve grafiksel yöntemlerin incelenmesi faydalı olabilir (Mecklin ve Mundfrom, 2005; Korkmaz vd., 2014).

Çok değişkenli normal dağılıma ilişkin ilk çalışmalar 1898'de F.Galton tarafından yapılmıştır. İki değişkenli istatistik yöntemlerden çok değişkenli istatistik yöntemlere geçişte önemli bir adım olmuştur (Biçkici, 2007). Mardia tarafından çok değişkenli basıklık ve çarpıklık ölçümlerine dayalı olarak Mardia'nın çok değişkenli normallik testi geliştirilmiştir. Cain ve arkadaşları, SAS, SPSS, R ve geliştirilmiş bir web uygulamasında bu testi incelemişlerdir (Mardia, 1970; Cain vd., 2017).

3.1.1. Normal Dağılıma Uygunluk Yöntemlerinin Karşılaştırmaları

Normal dağılıma uygunluk için kaynaklarda bazı yöntemler önerilmektedir. Araştırmacının tek yöntemle karar vermek yerine, örneklemin büyüklüğü ve

araştırmanın önemine göre konu ile ilgili önceki birikimlerinden faydalanarak birkaç yöntemi dikkate alarak karar vermesinin en doğru yöntem olacağı düşünülmektedir (Tarı, 2008).

- Sadece çarpıklık ve basıklık değerlerine göre değerlendirme yapan yöntem için farklı aralıklar önerilebilmektedir.
- Yukarıdakine benzer ancak çarpıklık ve basıklık değerlerinin kendi standart hatalarına bölündükten sonra sonucun belli aralıklarda olmasını öneren yöntemler vardır.
- Küçük örneklemler için Shapiro-Wilk, büyük örneklemler için Kolmogorov-Smirnov (Lilliefors) değerlerinin anlamlı yani 0,05'den büyük olması istenmektedir.

Her durumda da Shapiro-Wilk'e bakılmasını önerenler de vardır.

- Tek örnek Kolmogorov-Smirnov Testinin anlamlılık düzeylerine göre incelenmesi önerilmektedir. Anlamlılığı 0,05'den büyük olanların normal dağılıma uygun olacağı belirtilmektedir.
- Grafik yöntemlere bakılarak karar verilebileceği söylenmektedir.

Ancak bu yöntemlerin hepsi aynı sonucu vermeyebilirler. Hatta benzer olan ilk iki yöntemden hangisinin ve hangi aralığın kabul edileceğine dair kesin bir görüş birliği de bulunmamaktadır.

Hassas olan Shapiro-Wilk'e her durumda bakılması büyük örneklerde normal dağılıma uyan verilerin normalliğini reddetmeye neden olabilmektedir. Düzeltilmiş Kolmogorov-Smirnov (Lilliefors) test sonuçları genellikle Tek Örnek Kolmogorov-Smirnov Test sonuçlarından farklı çıkabilmektedir. Sadece grafiklere bakılarak yapılacak değerlendirme ise sadece gözleme dayalı olması nedeniyle şüphe uyandırabilmektedir (Demir vd., 2016; ankara.edu.tr 25.09.2019-web; kemaldoymus 25.09.2019-web).

Aşağıdaki başlıklarda tek değişkenli normalliği test eden bu yöntemler arasında karşılaştırmalar yapılmış olup, değişkenlerin ayrı ayrı tek değişkenli normalliği sağlayıp sağlamama durumlarına karar verilmiştir. Bu kararlar, aynı değişkenler kümesine Mardia'nın çok değişkenli normal dağılım testi sonucundaki çok değişkenli normal dağılıma uygunluklar ile karşılaştırılmıştır.

3.1.2. Tek ve Çok Değişkenli Normallik Şartını Sağlayan Veriler

Bir iş yerinde; müşteri temsilcisi, muhasebe ve makinacı olarak çalışanların çalışma alanları tecrübe, giyim tarzı ve bilgisayar bilgisi düzeylerine göre yapılan puanlama ile belirlendiği varsayılmaktadır. 245 gözlemden oluşan ve DATA-1 olarak adlandırılan verilerin normal dağılıma uygunlukları bu başlıkta incelenecek olup, ilerleyen bölümlerde de aynı adla ayrıca kullanılacaktır. DATA-1 verilerine ait çarpıklık ve basıklık ölçüleri Tablo 3.1'de görülmektedir.

Tablo 3.1. DATA-1 Tanımlayıcı İstatistikler

		İstatistik	**Std. Hata**
Tecrübe	Aritmetik ortalama	12,000	0,328
	Çarpıklık	0,000	0,156
	Basıklık	-0,581	0,310
Giyim Tarzı	Aritmetik ortalama	20,633	0,360
	Çarpıklık	-0,037	0,156
	Basıklık	-0,523	0,310
Bilgisayar Düzeyi	Aritmetik ortalama	9,167	0,247
	Çarpıklık	-0,017	0,156
	Basıklık	-0,432	0,310

Hesaplamalar sonucunda elde edilen çarpıklık ve basıklık değerlerinin tümü ±3 aralığında yer aldığından değişkenlerin tümünün tek değişkenli normalliğe uygun olduğu söylenebilir. Verilerin normal dağılması için kabul edilen çarpılık ve basıklık değerlerinin aralığını ±2 ve ±1 olarak kabul eden kaynaklar için de aynı yorum yapılabilir (Kalaycı, 2010; ankara.edu.tr 25.09.2019-web; kemaldoymus 25.09.2019-web). Normallik değerlendirmesi için çarpıklık ve basıklık değerlerinin kendi standart hatalarına bölünmesini öneren yöntem için aşağıdaki hesaplamalar yapılmıştır.

Tecrübe : $\frac{\text{Çarpıklık}}{\text{Çar.St.Hata}} = \frac{0,000}{0,156} = 0,00$ $\qquad \frac{\text{Basıklık}}{\text{Bas. St. Hata}} = \frac{-0,581}{0,310} = -1,90$

Giyim tarzı : $\frac{\text{Çarpıklık}}{\text{Çar. St. Hata}} = \frac{-0,037}{0,156} = -0,24$ $\qquad \frac{\text{Basıklık}}{\text{Bas. St. Hata}} = \frac{-0,523}{0,310} = -1,69$

Bil.düzeyi : $\frac{\text{Çarpıklık}}{\text{Çar. St. Hata}} = \frac{-0,017}{0,156} = -0,11$ $\qquad \frac{\text{Basıklık}}{\text{Bas. St. Hata}} = \frac{-0,432}{0,310} = -1,40$

Sonuçlar, 0,05 anlamlılık düzeyinde anlamsız çıktığından yani standartlaştırılmış değerler ±1,96 aralığı içinde olduğundan değişkenlerin tümünün normal dağıldığı söylenebilmektedir.

Normallik testi için kullanılan analitik testlerden Kolmogorov-Simirnov ve Shapiro-Wilk test istatistikleri Tablo 3.2'de verilmiştir.

Tablo 3.2. DATA-1 Kolmogorov-Simirnov ve Shapiro-Wilk Normallik Testleri

	Kolmogorov-Smirnov			**Shapiro-Wilk**		
	İstatistik	**df**	**Sig.**	**İstatistik**	**df**	**Sig.**
Tecrübe	0,039	245	0,200	0,989	245	0,070
Giyim Tarzı	0,051	245	0,200	0,992	245	0,196
Bilgisayar Düzeyi	0,052	245	0,200	0,990	245	0,074

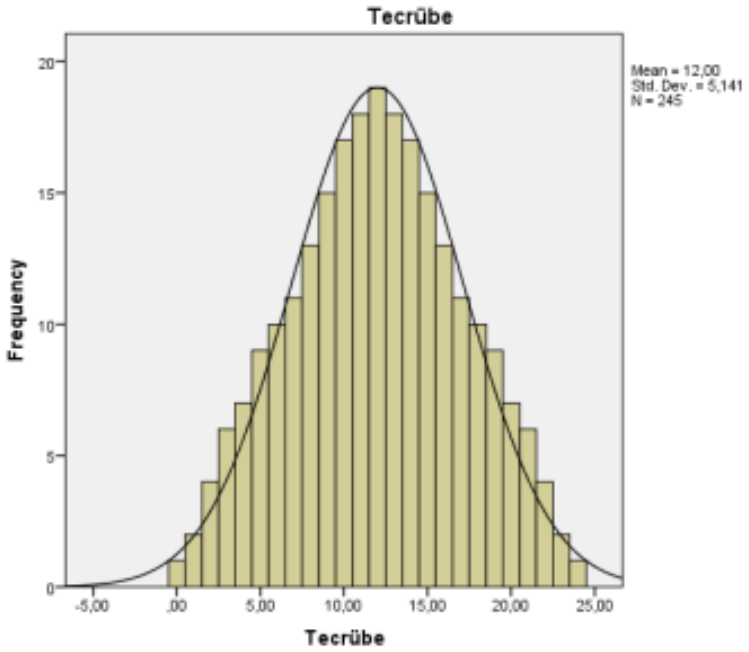
Bazı araştırmacılara göre 30'dan bazılarına göre 50'den az olan küçük örneklemler için önerilen ve daha hassas olan Shapiro-Wilk testine göre de Kolmogorov-Smirnov testine göre de her üç değişken normal dağılıma uygun bulunmuştur.

Tablo 3.3. DATA-1 Tek Örnek Kolmogorov-Simirnov Normallik Testi

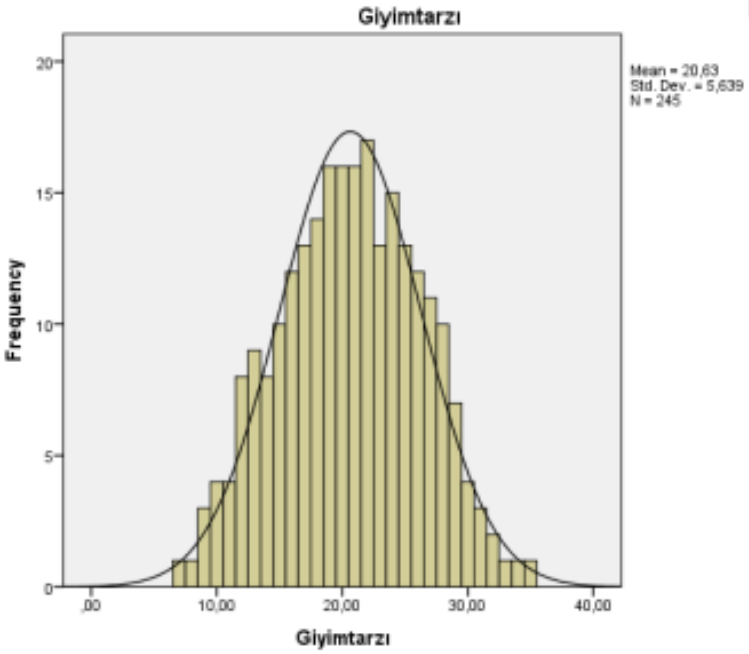
		Tecrübe	**Giyim Tarzı**	**Bilgisayar Düzeyi**
N		245	245	245
Normal Parametreler	Aritmetik ortalama	12,000	20,633	9,167
	Standart Sapma	5,141	5,639	3,871
En uç farklar	Mutlak	0,039	0,051	0,052
	Pozitif	0,039	0,039	0,052
	Negatif	-0,039	-0,051	-0,050
Kolmogorov-Smirnov Z		0,612	0,803	0,813
Asymp. Sig. (2-tailed)		0,848	0,539	0,523

Yukarıdaki Tablo 3.3 ile verilen Tek Örnek Kolmogorov-Smirnov testine göre üç değişkende normal dağılıma uygun bulunmuştur.

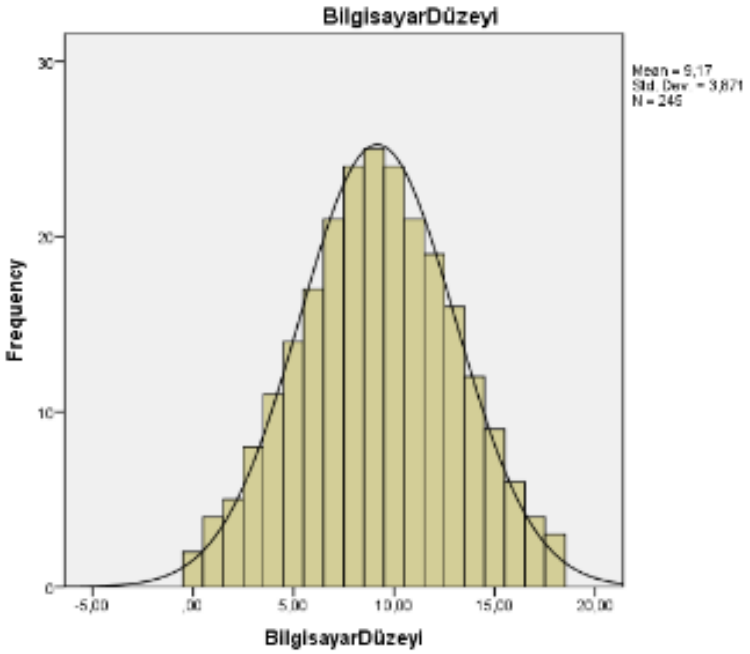
Bu verilere ait görsel testlerden aşağıdaki Şekil 3.1, Şekil 3.2, Şekil 3.3'de verilen histogram grafikleri de incelenebilir. Grafikler incelendiğinde değişkenlerin görsel olarak da normal dağılıma uygun olduğu söylenebilir. Kişisel olarak farklı grafik türleri de tercih edilebilir.



Şekil 3.1. DATA-1 Tecrübe Grafiği



Şekil 3.2. DATA-1 Giyim Tarzı Grafiği



Şekil 3.3. DATA-1 Bilgisayar Düzeyi Grafiği

Bu verideki değişkenler için yapılan normallik testlerine göre çarpıklık ve basıklık katsayıları veya bu değerlerin standart hatalarına bölümünden elde edilen sonuçlar incelenebilir. Diğer test istatistikleri ve grafikler incelendiğinde bu değişkenlerin ayrı ayrı normal dağılıma uygun oldukları görülebilir.

DATA-1'de yer alan değişkenlerin tek değişkenli normallik varsayımını sağladıkları söylenebilir. Ayrı ayrı tek değişkenli normal dağılıma uyan bu değişkenler, çoklu normal dağılıma uygun bulunmayabilirler. Bundan dolayı verilere çoklu normallik varsayımını gerektiren çok değişkenli istatistik yöntemlerinin uygulanabilmesi için bu varsayımın da test edilmesi gerekir.

Bu veriler için Mardia'nın çok değişkenli çarpıklık ve basıklık ölçüleri Tablo 3.4'de verilmiştir. Aşağıdaki tablo incelendiğinde anlamlı bulunan p değerine göre bu değişkenlerin çok değişkenli normallik varsayımını da sağladığı söylenebilir.

Tablo 3.4. DATA-1 Mardia'nın Çoklu Normallik Testi

	Hesaplanan Değerler	**Standartlaştırılmış Değerler**	**p değeri**
Çok Değişkenli Çarpıklık	0,223	9,111	0,522
Çok Değişkenli Basıklık	13,907	-1,561	0,118

3.1.3. Tek Değişkenli Normallik İçin Uygun Görünmeyen, Çok Değişkenli Normal Dağılıma Uyan Değişkenler

DATA-1 olarak oluşturulan simülasyon verilerinin tek değişkenli normalliğe uygunlukları sağlanarak, çok değişkenli normal dağılıma uygunlukları incelenmişti. Tek değişkenli normal dağılıma uymayan verilerin, çok değişkenli normal dağılım testinde verdiği sonucu görmek için, DATA-1 araştırmasına konu olan, bir iş yeri çalışanlarının departmanını belirlemeye yönelik puanlama çalışmasının DATA-2 olarak adlandırılan farklı bir simülasyonu incelenmiştir. Tek değişkenli normal dağılıma uymayan verilerin çok değişkenli normal dağılıma uygunlukları test edilmiştir. DATA-2 verilerine ait çarpıklık ve basıklık ölçüleri Tablo 3.5'de verilmiştir.

Tablo 3.5. DATA-2 Tanımlayıcı İstatistikler

		İstatistik	**Std. Hata**
Tecrübe	Aritmetik ortalama	15,641	0,309
	Çarpıklık	-0,420	0,156
	Basıklık	0,295	0,310
Giyim Tarzı	Aritmetik ortalama	20,682	0,349
	Çarpıklık	-0,092	0,156
	Basıklık	-0,650	0,310
Bilgisayar Düzeyi	Aritmetik ortalama	10,584	0,238
	Çarpıklık	0,053	0,156
	Basıklık	-0,314	0,310

Sadece bu tabloya göre değerlendirme yapılacak olursa çarpıklık ve basıklık değerlerinin tümü ±1 aralığında yer almakta olup, değişkenlerin tümünün tek değişkenli normalliğe uygun olduğu söylenebilir.

Çarpıklık ve basıklık değerlerinin kendi standart hatalarına bölünmesini öneren yöntem için aşağıdaki hesaplamalar yapılmıştır.

Tecrübe : $\frac{Çarpıklık}{Çar.St.Hata} = \frac{-0,420}{0,156} = -2,70$ $\frac{Basıklık}{Bas.St.Hata} = \frac{0,295}{0,310} = 1,00$

Giyim tarzı : $\frac{Çarpıklık}{Çar.St.Hata} = \frac{-0,092}{0,156} = -0,59$ $\frac{Basıklık}{Bas.St.Hata} = \frac{-0,650}{0,310} = -2,10$

Bilgisayar düzeyi : $\frac{Çarpıklık}{Çar.St.Hata} = \frac{0,053}{0,156} = 0,34$ $\frac{Basıklık}{Bas.St.Hata} = \frac{-0,314}{0,310} = -1,01$

Bu değerler normallik için önerilen ±1,96 aralığı içinde olmayıp -3 yaklaşan değerler bulunmaktadır.

Normallik testi için kullanılan analitik testlerden Kolmogorov-Simirnov ve Shapiro-Wilk test istatistikleri DATA-2 için Tablo 3.6'da verilmiştir.

Tablo 3.6. DATA-2 Kolmogorov-Simirnov ve Shapiro-Wilk Normallik Testleri

	Kolmogorov-Smirnov			**Shapiro-Wilk**		
	İstatistik	**df**	**Sig.**	**İstatistik**	**df**	**Sig.**
Tecrübe	0,085	245	0,000	0,984	245	0,009
Giyim Tarzı	0,071	245	0,005	0,987	245	0,023
Bilgisayar Düzeyi	0,073	245	0,003	0,988	245	0,038

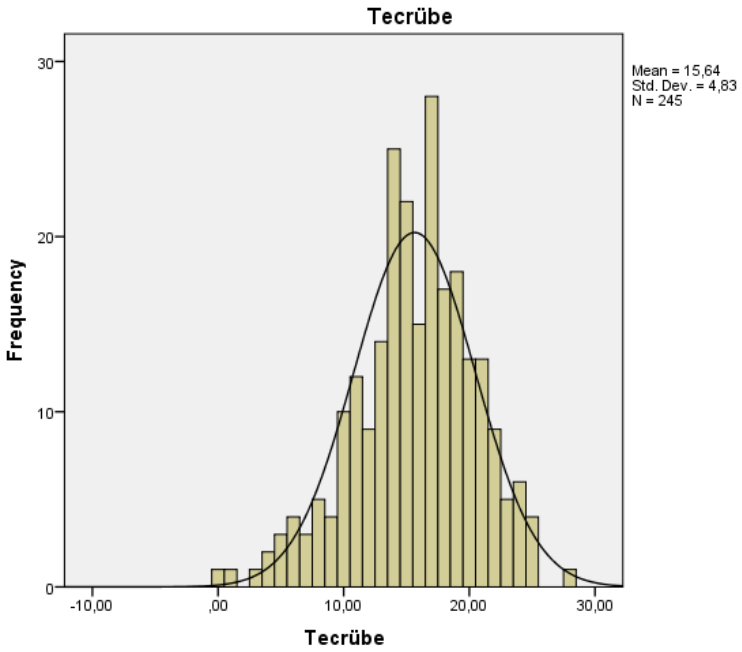
Tablo 3.6'ya göre hem Shapiro-Wilk hem de Kolmogorov-Smirnov (Lilliefors) testine göre her üç değişken de normal dağılıma uygun değildir.

Tablo 3.7. DATA-2 Tek Örnek Kolmogorov-Simirnov Normallik Testi

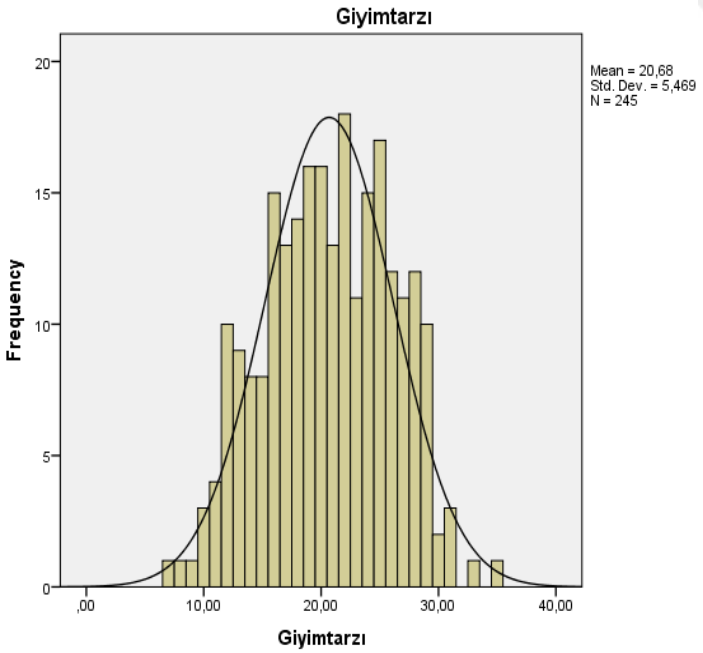
		Tecrübe	**Giyim Tarzı**	**Bilgisayar Düzeyi**
N		245	245	245
Normal Parametreler	Aritmetik ortalama	15,641	20,682	10,584
	Standart Sapma	4,830	5,469	3,721
En uç farklar	Mutlak	0,085	0,071	0,073
	Pozitif	0,038	0,049	0,073
	Negatif	-0,085	-0,071	-0,073
Kolmogorov-Smirnov Z		1,337	1,109	1,144
Asymp. Sig. (2-tailed)		0,056	0,171	0,146

Tablo 3.7 Tek Örnek Kolmogorov-Simirnov normallik testi sonuçlarına göre tecrübe değişkeni anlamlılık sınırına yakın olmak üzere üç değişkende normal dağılıma uygun bulunmuştur.

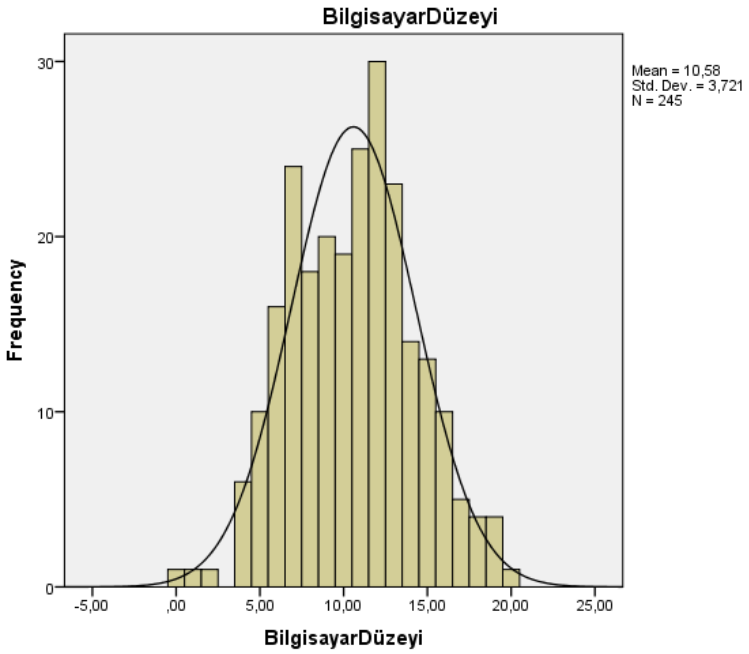
DATA-2 verilerine ait histogram grafikleri Şekil 3.4, Şekil 3.5 ve Şekil 3.6'da verilmiştir. Grafiklerin normal dağılım grafiğine tam olarak uyduğu söylenemeyebilir.



Şekil 3.4. DATA-2 Tecrübe Grafiği



Şekil 3.5. DATA-2 Giyim Tarzı Grafiği



Şekil 3.6. DATA-2 Bilgisayar Düzeyi Grafiği

Çarpıklık ve basıklık değerlerine göre değerlendirme yapan yönteme ve Tek Örnek Kolmogorov-Simirnov normallik testine göre değişkenlerden biri anlamlılık sınırına yakın da olsa normal dağılıma uygun bulunmuş fakat diğer yöntemlere göre ise normal dağılıma uygun bulunmadığından değişkenlerin genel olarak tek değişkenli normal dağılıma uygun olmadığı söylenebilir.

Bu veriler için değişkenler ayrı ayrı tek değişkenli normal dağılım açısından değerlendirildiğinde normal dağılıma uymadığı söylenebilecek iken Tablo 3.8'de Mardia'nın çok değişkenli normal dağılım testine göre değişkenler çoklu normal dağılıma uygun bulunmuştur.

Tablo 3.8. DATA-2 İçin Mardia'nın Çoklu Normallik Testi

	Hesaplanan Değerler	**Standartlaştırılmış Değerler**	**p değeri**
Çok Değişkenli Çarpıklık	0.307	12.551	0.249
Çok Değişkenli Basıklık	13.883	-1.595	0.111

3.1.4. Tek Değişkenli Normallik İçin Uygun Görünen, Çok Değişkenli Normal Dağılıma Uymayan Değişkenler

Bir spor okulunda öğrencilerin yatkın oldukları spor dalını belirlemek için açık atlama, smaç, krospas, kondüsyon ve sıçrama değişkenlerine ilişkin puanlamalar yapılarak, aldıkları puanlara göre öğrenciler jimnastik veya voleybola yönlendirilmekte oldukları varsayılan bir çalışma ele alınsın. 485 gözlemden oluşan ve DATA-3 olarak adlandırılan verilerin normal dağılıma uygunlukları bu bölümde incelenecek olup, ilerleyen bölümlerde de aynı adla ayrıca kullanılacaktır. Bu değişkenlerin tek ve çok değişkenli normal dağılıma uygunluk değerlendirmeleri aşağıdadır.

DATA-3 olarak verilerine ait çarpıklık ve basıklık ölçüleri Tablo 3.9'da yer almaktadır.

Tablo 3.9. DATA-3 Tanımlayıcı İstatistikler

		İstatistik	**Std. Hata**
Açık Atlama	Aritmetik ortalama	12,010	0,231
	Çarpıklık	0,000	0,111
	Basıklık	-0,539	0,221
Smaç	Aritmetik ortalama	11,920	0,229
	Çarpıklık	0,019	0,111
	Basıklık	-0,481	0,221
Krospas	Aritmetik ortalama	11,889	0,228
	Çarpıklık	0,023	0,111
	Basıklık	-0,456	0,221
Kondüsyon	Aritmetik ortalama	9,210	0,172
	Çarpıklık	-0,015	0,111
	Basıklık	-0,403	0,221
Sıçrama	Aritmetik ortalama	16,359	0,309
	Çarpıklık	0,063	0,111
	Basıklık	-0,433	0,221

Tablodaki çarpıklık ve basıklık değerlerinin tümü ±1 aralığında yer aldığından bu yönteme göre değişkenlerin tümünün tek değişkenli normalliğe uygun olduğu söylenebilir.

Normallik değerlendirmesi için çarpıklık ve basıklık değerlerinin kendi standart hatalarına bölünmesini öneren yöntemi değerlendirmek için aşağıdaki hesaplamalar yapılmıştır.

Açık atlama : $\frac{Çarpıklık}{Çar.St.Hata} = \frac{0,000}{0,111} = 0,00$ $\frac{Basıklık}{Bas.St.Hata} = \frac{-0,539}{0,221} = -2,44$

Smaç : $\frac{Çarpıklık}{Çar.St.Hata} = \frac{0,019}{0,111} = 0,17$ $\frac{Basıklık}{Bas.St.Hata} = \frac{-0,481}{0,221} = -2,18$

Krospas : $\frac{Çarpıklık}{Çar.St.Hata} = \frac{0,023}{0,111} = 0,21$ $\frac{Basıklık}{Bas.St.Hata} = \frac{-0,456}{0,221} = -2,06$

Kondüsyon : $\frac{Çarpıklık}{Çar.St.Hata} = \frac{-0,015}{0,111} = -0,14$ $\frac{Basıklık}{Bas.St.Hata} = \frac{-0,403}{0,221} = -1,82$

Sıçrama : $\frac{Çarpıklık}{Çar.St.Hata} = \frac{0,063}{0,111} = 0,57$ $\frac{Basıklık}{Bas.St.Hata} = \frac{-0,433}{0,221} = -1,95$

Değerlerin çoğunun ±1,96 aralığı içinde olduğu bazılarında biraz sapma olduğu söylenebilir.

Normallik testi için kullanılan analitik testlerden Kolmogorov-Simirnov ve Shapiro-Wilk test istatistikleri DATA-3 için Tablo 3.10'da verilmiştir.

Tablo 3.10. DATA-3 Kolmogorov-Simirnov ve Shapiro-Wilk Normallik Testleri

	Kolmogorov-Smirnov			**Shapiro-Wilk**		
	İstatistik	**df**	**Sig.**	**İstatistik**	**df**	**Sig.**
Açık Atlama	0,040	485	0,066	0,990	485	0,003
Smaç	0,040	485	0,056	0,992	485	0,009
Krospas	0,040	485	0,056	0,992	485	0,012
Kondüsyon	0,054	485	0,002	0,990	485	0,002
Sıçrama	0,037	485	0,142	0,994	485	0,055

Shapiro-Wilk istatistiğine göre bir değişken normalliği sağlamışken aynı tablonun Kolmogorov-Smirnov (Lilliefors) istatistiğine göre dört değişken normal dağılıma sahiptir.

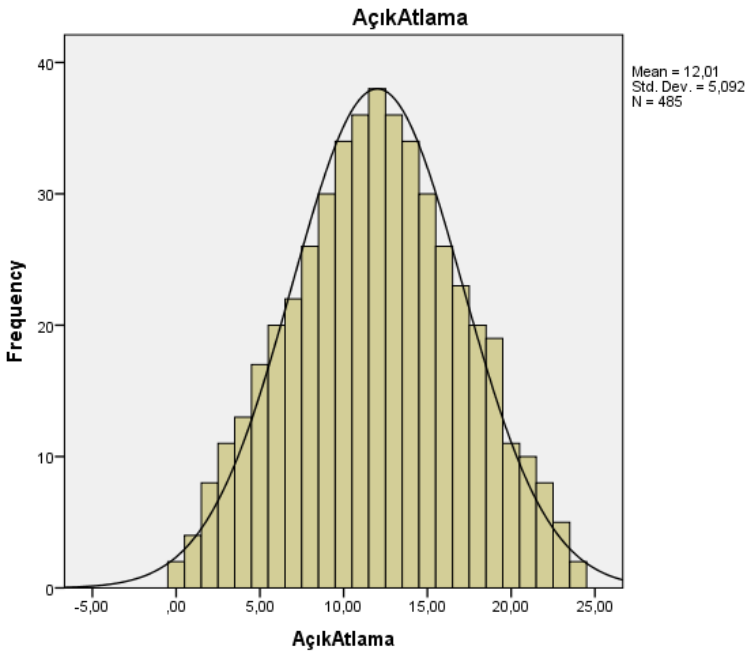
Tablo 3.11 ile Tek Örnek Kolmogorov-Simirnov yöntemi ile yapılan hesaplamalar verilmiştir.

Tablo 3.11. DATA-3 Tek Örnek Kolmogorov-Simirnov Normallik Testi

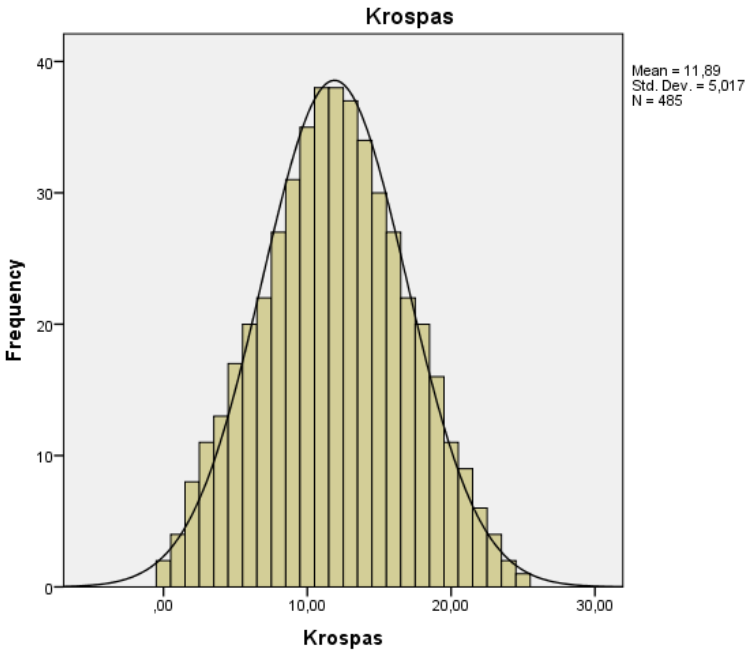
		Açık Atlama	Smaç	Krospas	Kondüsyon	Sıçrama
N		485	485	485	485	485
Normal Parametreler	Aritmetik ortalama	12,010	11,920	11,889	9,210	16,359
	Std. Sapma	5,092	5,041	5,017	3,790	6,801
En uç farklar	Mutlak	0,040	0,040	0,040	0,054	0,037
	Pozitif	0,039	0,040	0,040	0,054	0,037
	Negatif	-0,040	-0,039	-0,039	-0,051	-0,031
Kolmogorov-Smirnov Z		0,873	0,890	0,890	1,191	0,813
Asymp. Sig. (2-tailed)		0,431	0,406	0,407	0,117	0,524

Tabloya göre Kolmogorov-Smirnov istatistiğinde tüm değişkenler için normallik sağlanmıştır.

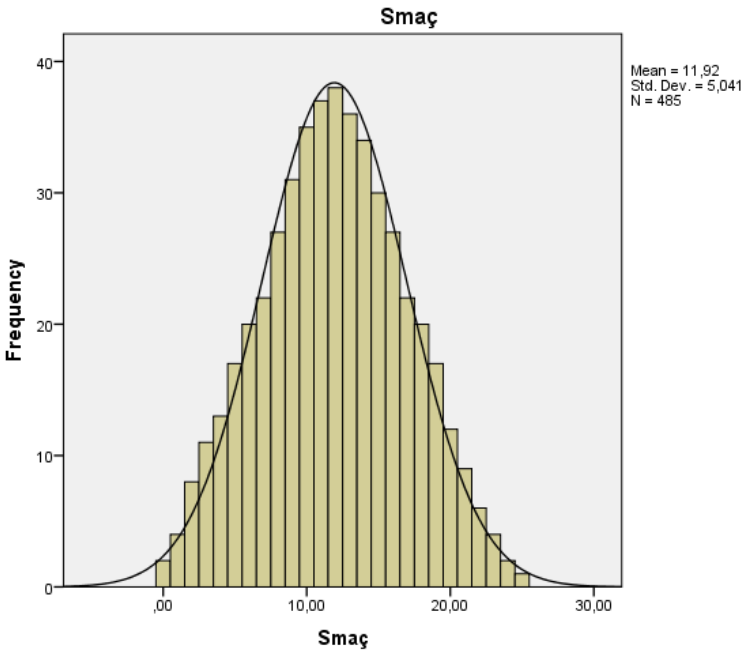
DATA-3 verilerine ait grafikler Şekil 3.7, Şekil 3.8, Şekil 3.9, Şekil 3.10 ve Şekil 3.11'de verilmiştir. Grafiklerin normal dağılım grafiğine benzer olduğu söylenebilir.



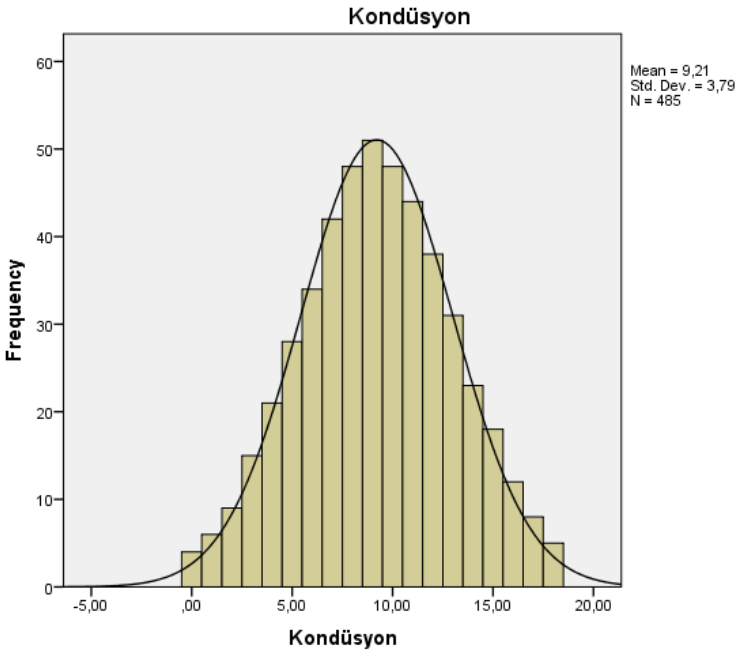
Şekil 3.7. DATA-3 Açık Atlama Grafiği



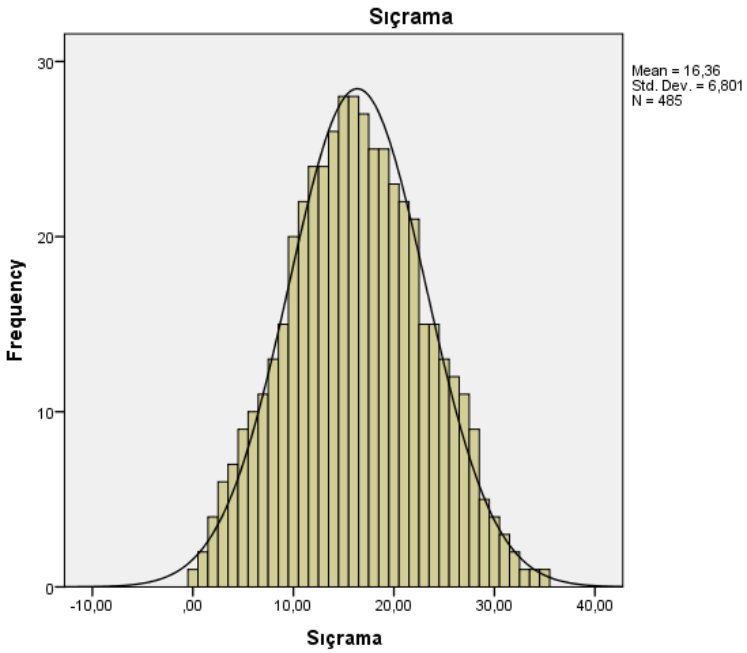
Şekil 3.8. DATA-3 Krospas Grafiği



Şekil 3.9. DATA-3 Smaç Grafiği



Şekil 3.10. DATA-3 Kondüsyon Grafiği



Şekil 3.11. DATA-3 Sıçrama Grafiği

Gözlem sayısının 485 olduğu, grafiklerin ve istatistiklerin birlikte değerlendirilmesi ile bu değişkenlerin ayrı ayrı normal dağılıma uygun olduğu kabul edilebilir. Ancak Tablo 3.12 ile verilen Mardia'nın normal dağılım testine göre değişkenler çoklu normal dağılıma uygun bulunmamıştır.

Tablo 3.12. DATA-3 İçin Mardia'nın Çoklu Normallik Testi

	Hesaplanan Değerler	**Standartlaştırılmış Değerler**	**p değeri**
Çok Değişkenli Çarpıklık	9.907	800.821	0.000
Çok Değişkenli Basıklık	56.637	28.477	0.000

3.1.5. Tek Değişkenli Normal Dağılım Testlerinin Hassasiyeti

Normallik değerlenmelerinde yukarıdaki karşılaştırmalar incelendiğinde bir değişkeni normal dağılıma uygun bulma aleyhine olan sıralamanın; Shapiro-Wilk istatistiği, Kolmogorov-Smirnov (Lilliefors) testi, Tek Örnek Kolmogorov-Smirnov Testinin Kolmogorov-Smirnov Z istatistiği olarak sıralanabilir.

Fakat Shapiro-Wilk istatistiğinde bazen normale oldukça yakın olan gözlemleri anlamsız olarak değerlendirebilme ihtimali, Kolmogorov-Smirnov istatistiğinde ise normalden uzak verilerin de normal kabul edilebilme ihtimali olabilir. Bu durum, yukarıdaki veri setleri dışında farklı değişkenler üzerinde aşamalı olarak tekrar incelenebilir. 100 futbolcunun başarı sıralamalarını yapmak için bu futbolcuların gol, isabetli pas ve asist sayılarına ait DATA-4 olarak adlandırılan bir çalışma değerlendirilmeye alınsın. Tek değişkenli normal dağılım testlerinin, değişkenleri normal dağılıma uygun bulma sıralamaları için aşağıdaki testler yapılmıştır.

DATA-4 verilerine ait çarpıklık ve basıklık ölçüleri Tablo 3.13'de verilmiştir.

Tablo 3.13. DATA-4 Tanımlayıcı İstatistikler

		İstatistik	**Std. Hata**
Gol	Aritmetik ortalama	2,50	0,168
	Çarpıklık	0,000	0,241
	Basıklık	-1,209	0,478
İsabetli Pas	Aritmetik ortalama	15,09	0,768
	Çarpıklık	-0,343	0,241
	Basıklık	-1,011	0,478
Asist	Aritmetik ortalama	4,70	0,372
	Çarpıklık	0,678	0,241
	Basıklık	-0,598	0,478

Çarpıklık ve basıklık değerlerinin tümü ±2 aralığında olup hatta ±1 aralığına oldukça yakındır. Sadece bu değerlendirmeye göre değişkenlerin tümünün tek değişkenli normalliğe uygun olduğu söylenebilir.

Çarpıklık ve basıklık değerlerinin kendi standart hatalarına bölünmesini öneren yöntem için aşağıdaki hesaplamalar yapılmıştır.

Gol : $\frac{Çarpıklık}{Çar.St.Hata} = \frac{0,000}{0,241} = 0,00$ $\frac{Basıklık}{Bas.St.Hata} = \frac{-1,209}{0,478} = -2,53$

İsabetli pas : $\frac{Çarpıklık}{Çar.St.Hata} = \frac{-0,343}{0,241} = -1,42$ $\frac{Basıklık}{Bas.St.Hata} = \frac{-1,011}{0,478} = -2,12$

Asist : $\frac{Çarpıklık}{Çar.St.Hata} = \frac{0,678}{0,241} = 2,81$ $\frac{Basıklık}{Bas.St.Hata} = \frac{-0,598}{0,478} = -1,25$

Bu değerler normallik için önerilen ±1,96 aralığı içinde olmayıp ±3 yaklaşan değerler bulunmaktadır. Yukarıdaki yöntem ile tüm değişkenler normalmiş gibi değerlendirilecek iken bu yöntemde normallikten sapmalar olduğu söylenebilir.

Normallik testi için kullanılan analitik testlerden Kolmogorov-Simirnov ve Shapiro-Wilk test istatistikleri DATA-4 için Tablo 3.14'de verilmiştir.

Tablo 3.14. DATA-4 Kolmogorov-Simirnov ve Shapiro-Wilk Normallik Testleri

	Kolmogorov-Smirnov			Shapiro-Wilk		
	İstatistik	**df**	**Sig.**	**İstatistik**	**df**	**Sig.**
Gol	0,133	100	0,000	0,912	100	0,000
İsabetli Pas	0,088	100	0,053	0,948	100	0,001
Asist	0,148	100	0,000	0,915	100	0,000

Tabloya göre Shapiro-Wilk istatistiğinde hiçbir değişkenin normalliği anlamlı bulunmamıştır. Kolmogorov-Simirnov (Lilliefors) testine göre sadece isabetli pas değişkeni normallik varsayımını sağlamıştır. Bu değişkenin de anlamlılığın 0,05 sınır değerini çok az bir farkla geçtiği görülmektedir.

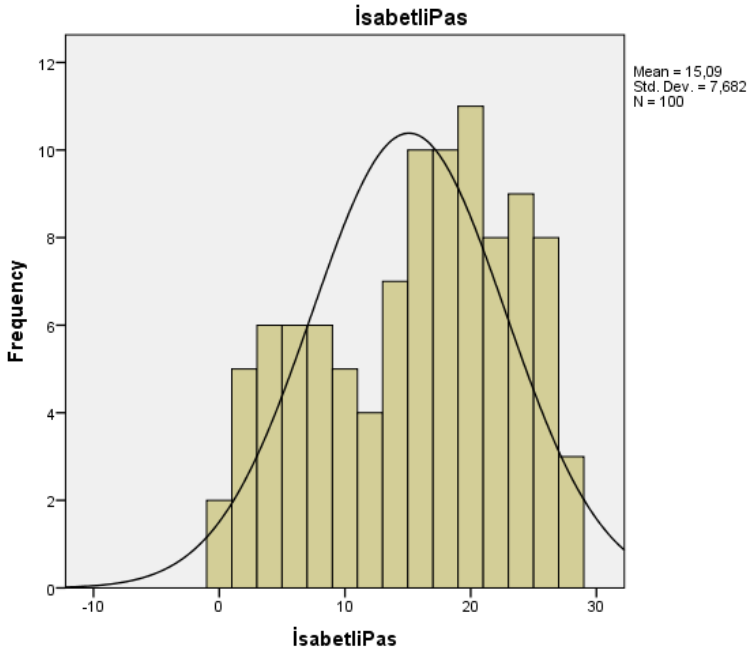
Tek Örnek Kolmogorov-Simirnov normallik testi Tablo 3.15 ile verilmiştir.

Tablo 3.15. DATA-4 Tek Örnek Kolmogorov-Simirnov Normallik Testi

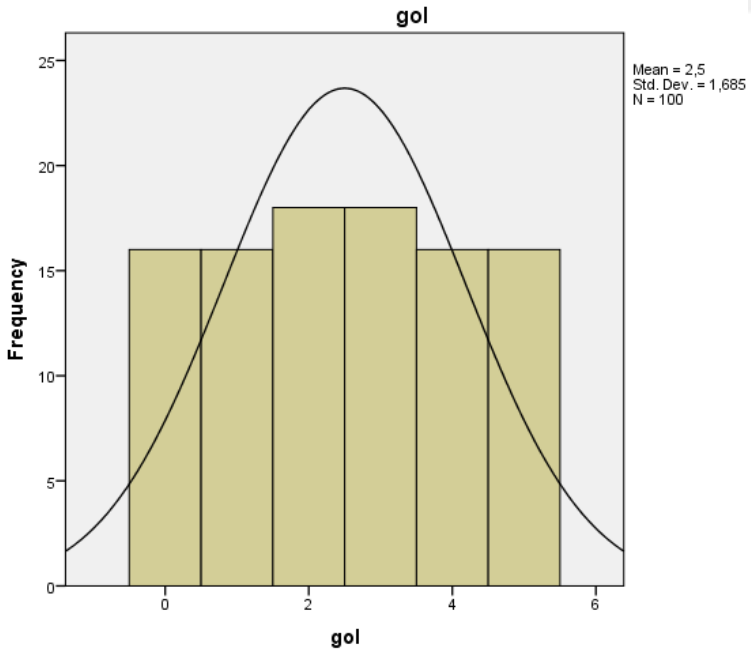
		Gol	**İsabetli Pas**	**Asist**
N		100	100	100
Normal Parametreler	Aritmetik ortalama	2,500	15,090	4,700
	Standart Sapma	1,685	7,682	3,724
En uç farklar	Mutlak	0,133	0,088	0,148
	Pozitif	0,133	0,074	0,148
	Negatif	-0,133	-0,088	-0,103
Kolmogorov-Smirnov Z		1,334	0,882	1,479
Asymp. Sig. (2-tailed)		0,057	0,418	0,025

Tablo 3.14'de Kolmogorov-Simirnov (Lilliefors) testine göre isabetli pas değişkeninin anlamlılığı artmış, gol değişkeni de normal dağılım için anlamlı olmuş, böylece iki değişkenin normal dağılıma uygun bulunduğu kabul edilmiştir. Tek Örnek Kolmogorov-Simirnov normallik testi, hem Kolmogorov-Simirnov (Lilliefors) hem de Shapiro-Wilk normallik testlerinin normal kabul etmediği değişkenleri normal kabul etmiştir.

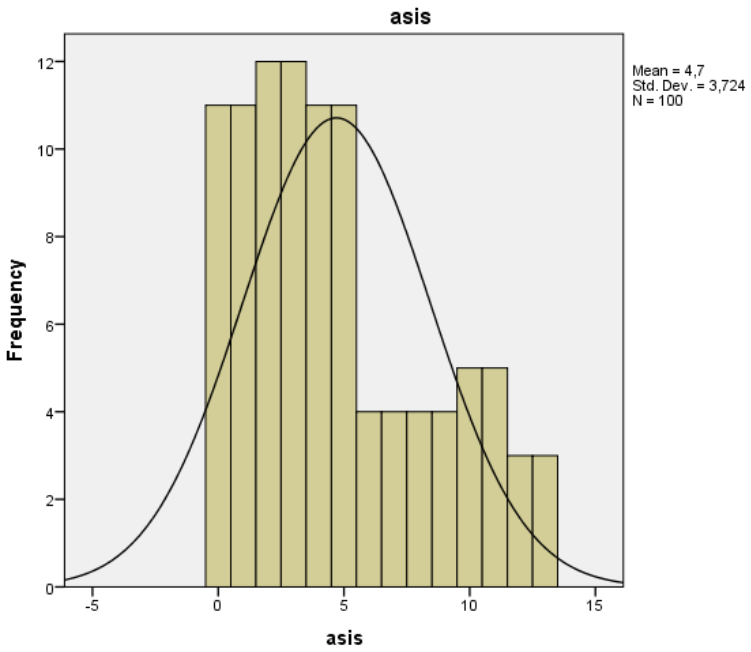
DATA-4 verilerine ait histogram grafikleri Şekil 3.12, Şekil 3.13 ve Şekil 3.14 verilmiştir.



Şekil 3.12. DATA-4 İsabetli Pas Grafiği



Şekil 3.13. DATA-4 Gol Grafiği



Şekil 3.14. DATA-4 Asist Grafiği

Grafikler incelendiğinde her üçünün de tam olarak normale yakın olmadığı söylenebilir. Shapiro-Wilk testi de bu şekilde sonuç vermişti. Kolmogorov-Simirnov (Lilliefors) testi sadece isabetli pas değişkeninin normalliğini anlamlı bulmuştu. Grafiklerden de normal dağılıma en yakın değişken olarak görülmektedir. Kolmogorov-Simirnov testinde isabetli pas değişkeninin anlamlılık düzeyi, Kolmogorov-Simirnov (Lilliefors)'a göre artmakta ve gol değişkeni de anlamlı kabul edilmektedir.

3.2. Varyans-Kovaryans Matrislerinin Eşitliği

Tek değişkenli istatistik yöntemler için varyans homojenliği varsayımında, varyansın bütün hücreler için eşit olması gerekir. Bu durum varyansların homojenliği olarak da ifade edilir. Çok değişkenli istatistik analizlerde verilerde birden fazla değişken olduğu için değişkenler arasındaki ilişki varyans-kovaryans matrisleri ile ifade edileceğinden tek değişkenli yöntemlerdeki varyansların homojenliği varsayımı, varyans-kovaryans matrislerinin eşitliği varsayımına dönüşür. Varyans-Kovaryans matrislerinin eşitliği Box-M testi ile test edilir. Ancak bu test normallikten sapmalara karşı çok duyarlı olduğundan çoklu normallik varsayımı sağlandıktan sonra bu testin yapılması gerekmektedir (Alpar, 2017). Box-M testinin normalliğe karşı aşırı duyarlı

olması bu testin sonuçlarının göz ardı edilmesine neden olmaktadır. Örneklem büyüklüklerinin durumlarına göre anlamlılık için Wilks Lamda, Hotelling, Roy'un en büyük kök veya Pillai ölçütü tercih edilebilir (Tabachnick ve Fidell, 2015).

3.3. Çoklu Bağlantı

Bağımsız değişkenler arasında ilişki olması durumuna çoklu bağlantı denilmektedir. Değişkenlerden bazıları başka değişkenler tarafından açıklanıyor olabilir, bu durumda hem mantıksal hem de istatistiksel problemlerle karşılaşılabilir. İstatistiksel problemler değişkenler arasındaki ilişki 0,90 veya daha yüksek olduğu zaman ortaya çıkar. Çoklu bağlantı matrisin tersini bulmayı zorlaştırırken, tekillik tamamen imkansız hale getirir. Bu problemi ortadan kaldırmak için değişkenler tekrar gözden geçirilip gerekli düzeltmeler yapılabilir. Hatta bazı değişkenler analizden çıkarılabilir (Tabachnick ve Fidell, 2015).

3.4. Doğrusallık

Korelasyon katsayılarına dayanan çok değişkenli istatistik tekniklerin varsayımlarından biri de doğrusallıktır. Bazı çok değişkenli istatistiksel yöntemler tüm bağımlı değişken çiftleri, tüm kovaryant çiftleri ve tüm bağımlı-kovaryant çiftlerinin doğrusal olması varsayımına dayanır. Doğrusallıktan sapmalar istatistiksel testin gücünü azaltır (Tabachnick ve Fidell, 2015).

3.5. Aşırı Değerler

Ortalamayı ciddi anlamda etkileyecek bir gözlem, değişkenliği de gereğinden fazla artıracaktır. Bu ise sonuçları etkileyerek yanlış yorumlara neden olacaktır. MANOVA, diskriminant analizi gibi grup ortalamaları ile çalışan yöntemler düşünüldüğünde konunun önemi daha iyi anlaşılacaktır. Aşırı değerler hem I. tip hem de II. tip hatalara yol açabilirler. Aşırı değerlerin olduğu verilerde genellemenin yapılması, sonuçların yanlış çıkmasına neden olabilir (Tabachnick ve Fidell, 2015). Bu nedenle her değişken grubu, çalışılan yöntem sınıflama yöntemi ise her grup, ayrı ayrı tek ve çok değişkenli olarak aykırı değerler açısından incelenmelidir.

Aşırı değerler değişkenlerin histogram, kutu-çizgi, dal-yaprak, ikişerli saçılım, kare Mahalanobis uzaklığı hesaplanarak elde edilen ki-kare saçılım, Chernoff yüzleri, ikon grafikleri gibi görsel testlerden veya değişkenlerin çok değişkenli normal dağılım göstermesi durumunda Barnett-Lewis tablolarından, genelleştirilmiş varyans oranı gibi analitik testlerden faydalanılarak belirlenebilir (Alpar, 2017).

4. KORELASYON KATSAYISI KARŞILAŞTIRMALARI

Çok değişkenli istatistik yöntemlerinin temeli değişkenler arasındaki ilişkilerdir. Değişkenler arasındaki ilişkiler ise kovaryans ve korelasyon katsayılarına dayanır.

Çoklu R, kanonik R, çok yönlü frekans analizi, çok düzeyli modelleme gibi değişkenler arasındaki ilişkinin derecesini ölçen çok değişkenli istatistiksel yöntemlerin kullanılmasında değişkenler arasında ilişki olması beklenir. MANOVA, MANCOVA gibi grup farklılıklarının anlamlılığının ölçüldüğü, çok değişkenli istatistiksel yöntemler kullanılırken gruplar içi değişkenlerin ilişkili olması gerekmektedir. Temel bileşenler analizi, Faktör Analizi, Yapısal Eşitlik Modellemesi gibi yapının incelendiği çok değişkenli istatistiksel yöntemlerde benzer gözlemler/değişkenler arasında ilişki olması istenir (Tabachnick ve Fidell, 2015).

Korelasyon analizi yapılırken değişkenler en az eşit aralıklı ölçeğe sahip ve normal dağılım gösteriyorsa Pearson korelasyon katsayısı kullanılır. Aynı değişken ölçeğinde veriler normal dağılıma uymuyorsa Spearman korelasyon katsayısı kullanılır. En çok ordinal ölçek türüne sahip verilerin korelasyonunu araştırmak için Kendall's tau istatistiğinden yararlanılır (anadolu.edu.tr 12.06.2019-web; Kalaycı, 2010; Kılıç, 2012). Bu bölümde Pearson, Spearman, Kendall's tau ilişki katsayıları, üçüncü bölümde normal dağılıma uygunlukları incelenen veriler kullanılarak karşılaştırılmıştır.

4.1. Normal Dağılıma Uyan Veriler İçin Korelasyon Katsayıları Karşılaştırması

Üçüncü bölümde normal dağılım testlerinin karşılaştırılmasında DATA-1 olarak verilen normal dağılıma uygun verilerin Pearson, Spearman ve Kendall's tau korelasyon katsayısı karşılaştırmaları Tablo 4.1 ve Tablo 4.2'de verilmiştir.

Tablo 4.1. DATA-1 Pearson Korelasyon Katsayıları

		Tecrübe	**Giyim Tarzı**	**Bilgisayar Düzeyi**
Tecrübe	Pearson Korelasyon	1	0,100	-0,060
	Sig. (2-tailed)		0,118	0,349
Giyim Tarzı	Pearson Korelasyon	0,100	1	0,048
	Sig. (2-tailed)	0,118		0,457
Bilgisayar Düzeyi	Pearson Korelasyon	-0,060	0,048	1
	Sig. (2-tailed)	0,349	0,457	

Tablo 4.2. DATA-1 Spearman ve Kendall Korelasyon Katsayıları

			Tecrübe	Giyim Tarzı	Bilgisayar Düzeyi
Kendall's tau_b	Tecrübe	Korelasyon Katsayısı	1,000	0,084	-0,057
		Sig. (2-tailed)	.	0,061	0,212
	Giyim Tarzı	Korelasyon Katsayısı	0,084	1,000	0,057
		Sig. (2-tailed)	0,061	.	0,211
	Bilgisayar Düzeyi	Korelasyon Katsayısı	-0,057	0,057	1,000
		Sig. (2-tailed)	0,212	0,211	.
Spearman's rho	Tecrübe	Korelasyon Katsayısı	1,000	0,117	-0,082
		Sig. (2-tailed)	.	0,068	0,202
	Giyim Tarzı	Korelasyon Katsayısı	0,117	1,000	0,074
		Sig. (2-tailed)	0,068	.	0,246
	Bilgisayar Düzeyi	Korelasyon Katsayısı	-0,082	0,074	1,000
		Sig. (2-tailed)	0,202	0,246	.

Tablo 4.1 ve Tablo 4.2 incelendiğinde normal dağılıma uyan bu veriler için her üç yöntem de benzer olarak aralarında korelasyon olan değişken olmadığına karar vermiştir.

Aşağıdaki tablolar da ise daha çok değişken içeren DATA-3 ile verilen normal dağılıma sahip değişkenler arasındaki korelasyon katsayıları karşılaştırılmıştır.

Tablo 4.3. DATA-3 Verileri İçin Pearson Korelasyon Katsayıları

		Açık Atlama	Smaç	Krospas	Kondüsyon	Sıçrama
Açık Atlama	Pearson Korelasyon	1	-0,998	-0,937	0,014	0,948
	Sig. (2-tailed)		0,000	0,000	0,755	0,000
Smaç	Pearson Korelasyon	-0,998	1	0,938	-0,013	-0,947
	Sig. (2-tailed)	0,000		0,000	0,770	0,000
Krospas	Pearson Korelasyon	-0,937	0,938	1	-0,020	-0,929
	Sig. (2-tailed)	0,000	0,000		0,654	0,000
Kondüsyon	Pearson Korelasyon	0,014	-0,013	-0,020	1	0,017
	Sig. (2-tailed)	0,755	0,770	0,654		0,714
Sıçrama	Pearson Korelasyon	0,948	-0,947	-0,929	0,017	1
	Sig. (2-tailed)	0,000	0,000	0,000	0,714	

Tablo 4.4. DATA-3 Verileri İçin Spearman ve Kendall Korelasyon Katsayıları

			Açık Atlama	Smaç	Krospas	Kondüsyon	Sıçrama
Kendall's tau_b	Açık Atlama	Korelasyon Katsayısı	1	-0,992	-0,847	0,012	0,856
		Sig. (2-tailed)		0,000	0,000	0,719	0,000
	Smaç	Korelasyon Katsayısı	-0,992	1	0,853	-0,012	-0,857
		Sig.(2-tailed)	0,000		0,000	0,713	0,000
	Krospas	Korelasyon Katsayısı	-0,847	0,853	1	-0,016	-0,864
		Sig.(2-tailed)	0,000	0,000		0,619	0,000
	Kondüsyon	Korelasyon Katsayısı	0,012	-0,012	-0,016	1	0,011
		Sig. (2-tailed)	0,719	0,713	0,619		0,731
	Sıçrama	Korelasyon Katsayısı	0,856	-0,857	-0,864	0,011	1
		Sig. (2-tailed	0,000	0,000	0,000	0,731	
Spearman's rho	Açık Atlama	Korelasyon Katsayısı	1	-0,999	-0,941	0,017	0,933
		Sig.(2-tailed)		0,000	0,000	0,703	0,000
	Smaç	Korelasyon Katsayısı	-0,999	1	0,942	-0,018	-0,933
		Sig.(2-tailed)	0,000		0,000	0,694	0,000
	Krospas	Korelasyon Katsayısı	-0,941	0,942	1	-0,019	-0,947
		Sig.(2-tailed)	0,000	0,000		0,682	0,000
	Kondüsyon	Korelasyon Katsayısı	0,017	-0,018	-0,019	1	0,017
		Sig.(2-tailed)	0,703	0,694	0,682		0,711
	Sıçrama	Korelasyon Katsayısı	0,933	-0,933	-0,947	0,017	1
		Sig.(2-tailed)	0,000	0,000	0,000	0,711	

Tablo 4.3 ve Tablo 4.4 değerlendirildiğinde korelasyon katsayılarının karşılaştırıldığı yöntemlerin üçü de aynı değişkenler arasında aynı yönlü ilişkiler bulmuştur. Ayrıca ilişki derecelerinin de birbirlerine yakın oldukları görülebilir. Bununla birlikte Pearson korelasyon katsayıları ile Spearman korelasyon katsayılarının

birbirlerine oldukça yakın oldukları da görülebilir. Sonuç olarak normal dağılıma uygun DATA-3 verileri için Pearson, Spearman's rho ve Kendall's tau_b korelasyon katsayıları yakın değerler vermiştir.

4.2. Normal Dağılıma Uymayan Verilerde Korelasyon Katsayısı Karşılaştırmaları

Üçüncü bölümde normalliği test edilen ve değişkenleri normal dağılıma uymadığı varsayılan DATA-2 verilerinin korelasyon analizi sonuçları Tablo 4.5 ve Tablo 4.6'da verilmiştir.

Tablo 4.5. DATA-2 Spearman ve Kendall Korelasyon Katsayıları

			Tecrübe	**Giyim Tarzı**	**Bilgisayar Düzeyi**
Kendall's tau_b	**Tecrübe**	Korelasyon Katsayısı	1	-0,052	0,059
		Sig. (2-tailed)	.	0,253	0,197
	Giyim Tarzı	Korelasyon Katsayısı	-0,052	1	-0,137
		Sig. (2-tailed)	0,253	.	0,002
	Bilgisayar Düzeyi	Korelasyon Katsayısı	0,059	-0,137	1
		Sig. (2-tailed)	0,197	0,002	.
Spearman's rho	**Tecrübe**	Korelasyon Katsayısı	1	-0,072	0,084
		Sig. (2-tailed)	.	0,260	0,189
	Giyim Tarzı	Korelasyon Katsayısı	-0,072	1	-0,198
		Sig. (2-tailed)	0,260	.	0,002
	Bilgisayar Düzeyi	Korelasyon Katsayısı	0,084	-0,198	1
		Sig. (2-tailed)	0,189	0,002	.

Tablo 4.6. DATA-2 Pearson Korelasyon Katsayıları

		Tecrübe	**Giyim Tarzı**	**Bilgisayar Düzeyi**
Tecrübe	Pearson Korelasyon	1	-0,071	0,079
	Sig. (2-tailed)		0,267	0,217
Giyim Tarzı	Pearson Korelasyon	-0,071	1	-0,236
	Sig. (2-tailed)	0,267		0,000
Bilgisayar Düzeyi	Pearson Korelasyon	0,079	-0,236	1
	Sig. (2-tailed)	0,217	0,000	

Veriler normal dağılım şartını sağlamadığı halde, normal dağılım şartı aranan Pearson korelasyon katsayısı dahil olmak üzere her üç korelasyon katsayısı da yaklaşık değerler almıştır. Her üç yöntemde giyim tarzı değişkeni ile bilgisayar düzeyi değişkenleri arasında güçlü olmayan korelasyon belirlemiştir.

4.3. Kategorik Değişkenli Verilerin Analizinde Korelasyon Karşılaştırmaları

Yukarıda korelasyon katsayısı karşılaştırmaları yapılan DATA-1, DATA-2 ve DATA-3 verilerinde değişkenler nicel verilerden oluşmaktaydı. Sigara içmeyi etkileyen faktörlerin araştırıldığı, DATA-5 olarak adlandırılan hem nitel hem de nicel değişkenler içeren bir çalışmanın değişkenleri için korelasyon katsayısı karşılaştırmaları yapılmış ve sonuçları Tablo 4.7, Tablo 4.8 ve Tablo 4.9'da verilmiştir.

Tablo 4.7. DATA-5 Pearson Korelasyon Katsayıları

		Sigara İçme	**Kanser Yaptığına İnanma**	**Cinsiyet**	**İnanca Aykırı Bulma**	**Kardeşin İçmesi**	**Gelir**	**Yaş**	**Medeni Hal**
Sigara İçme	Pearson Korelas.	1	-0,199	0,201	-0,160	0,163	-0,049	0,050	-0,104
	Sig. (2-Tailed)		0,003	0,003	0,018	0,016	0,473	0,465	0,128
Kanser Yaptığı İnanma	Pearson Korelas.	-0,199	1	0,041	0,083	-0,011	0,030	0,074	-0,104
	Sig. (2-Tailed)	0,003		0,548	0,225	0,870	0,657	0,275	0,125
Cinsiyet	Pearson Korelas.	0,201	0,041	1	0,198	0,105	-0,251	0,182	-0,169
	Sig. (2-Tailed)	0,003	0,548		0,003	0,123	0,000	0,007	0,013
İnanca Aykırı Bulma	Pearson Korelas.	-0,160	0,083	0,198	1	-0,024	-0,128	-0,001	-0,123
	Sig. (2-Tailed)	0,018	0,225	0,003		0,725	0,060	0,990	0,070
Kardeş. İçmesi	Pearson Korelas.	0,163	-0,011	0,105	-0,024	1	0,009	0,072	-0,122
	Sig. (2-Tailed)	0,016	0,870	0,123	0,725		0,890	0,289	0,073
Gelir	Pearson Korelas.	-0,049	0,030	-0,251	-0,128	0,009	1	0,171	-0,216
	Sig. (2-Tailed)	0,473	0,657	0,000	0,060	0,890		0,012	0,001
Yaş	Pearson Korelas.	0,050	0,074	0,182	-0,001	0,072	0,171	1	-0,305
	Sig. (2-Tailed)	0,465	0,275	0,007	0,990	0,289	0,012		0,000
Medeni Hal	Pearson Korelas.	-0,104	-0,104	-0,169	-0,123	-0,122	-0,216	-0,305	1
	Sig. (2-Tailed)	0,128	0,125	0,013	0,070	0,073	0,001	0,000	

Tablo 4.8. DATA-5 Kendall's tau_b Korelasyon Katsayıları

Kendall's tau_b									
		Sigara İçme	**Kanser Yaptığına İnanma**	**Cinsiyet**	**İnanca Aykırı Bulma**	**Kardeşin İçmesi**	**Gelir**	**Yas**	**Medeni Hal**
Sigara İçme	Korelasyon Katsayısı	1	-0,199	0,201	-0,160	0,163	-0,053	0,042	-0,104
	Sig. (2-tailed)	.	0,003	0,003	0,019	0,017	0,350	0,460	0,128
Kanser Yapt. İnanma	Korelasyon Katsayısı	-0,199	1	0,041	0,083	-0,011	0,071	0,058	-0,104
	Sig. (2-tailed)	0,003	.	0,546	0,225	0,870	0,216	0,307	0,125
Cinsiyet	Korelasyon Katsayısı	0,201	0,041	1	0,198	0,105	-0,193	0,158	-0,169
	Sig. (2-tailed)	0,003	0,546	.	0,004	0,123	0,001	0,005	0,013
İnanca Aykırı Bulma	Korelasyon Katsayısı	-0,160	0,083	0,198	1	-0,024	-0,073	0,009	-0,123
	Sig. (2-tailed)	0,019	0,225	0,004	.	0,724	0,197	0,872	0,070
Kardeş. İçmesi	Korelasyon Katsayısı	0,163	-0,011	0,105	-0,024	1	0,012	0,069	-0,122
	Sig. (2-tailed)	0,017	0,870	0,123	0,724	.	0,837	0,225	0,073
Gelir	Korelasyon Katsayısı	-0,053	0,071	-0,193	-0,073	0,012	1	0,135	-0,191
	Sig. (2-tailed)	0,350	0,216	0,001	0,197	0,837	.	0,004	0,001
Yaş	Korelasyon Katsayısı	0,042	0,058	0,158	0,009	0,069	0,135	1	-0,273
	Sig. (2-tailed)	0,460	0,307	0,005	0,872	0,225	0,004	.	0,000
Medeni Hal	Korelasyon Katsayısı	-0,104	-0,104	-0,169	-0,123	-0,122	-0,191	-0,273	1
	Sig. (2-tailed)	0,128	0,125	0,013	0,070	0,073	0,001	0,000	.

Tablo 4.9. DATA-5 Spearman's rho Korelasyon Katsayıları

Spearman's rho									
		Sigara İçme	**Kanser Yaptığına İnanma**	**Cinsiyet**	**Inanca Aykırı Bulma**	**Kardeşin İçmesi**	**Gelir**	**Yas**	**Medeni Hal**
Sigara İçme	Korelasyon Katsayısı	1	-0,199	0,201	-0,160	0,163	-0,064	0,050	-0,104
	Sig. (2-tailed)	.	0,003	0,003	0,018	0,016	0,351	0,461	0,128
Kanser Yapt. İnanma	Korelasyon Katsayısı	-0,199	1	0,041	0,083	-0,011	0,084	0,070	-0,104
	Sig. (2-tailed)	0,003	.	0,548	0,225	0,870	0,216	0,308	0,125
Cinsiyet	Korelasyon Katsayısı	0,201	0,041	1	0,198	0,105	-0,231	0,190	-0,169
	Sig. (2-tailed)	0,003	0,548	.	0,003	0,123	0,001	0,005	0,013
İnanca Aykırı Bulma	Korelasyon Katsayısı	-0,160	0,083	0,198	1	-0,024	-0,088	0,011	-0,123
	Sig. (2-tailed)	0,018	0,225	0,003	.	0,725	0,198	0,872	0,070
Kardeşin İçmesi	Korelasyon Katsayısı	0,163	-0,011	0,105	-0,024	1	0,014	0,083	-0,122
	Sig. (2-tailed)	0,016	0,870	0,123	0,725	.	0,838	0,225	0,073
Gelir	Korelasyon Katsayısı	-0,064	0,084	-0,231	-0,088	0,014	1	0,191	-0,228
	Sig. (2-tailed)	0,351	0,216	0,001	0,198	0,838	.	0,005	0,001
Yaş	Korelasyon Katsayısı	0,050	0,070	0,190	0,011	0,083	0,191	1	-0,328
	Sig. (2-tailed)	0,461	0,308	0,005	0,872	0,225	0,005	.	0,000
Medeni Hal	Korelasyon Katsayısı	-0,104	-0,104	-0,169	-0,123	-0,122	-0,228	-0,328	1
	Sig. (2-tailed)	0,128	0,125	0,013	0,070	0,073	0,001	0,000	.

Nitel ve nicel verilerin bir arada bulunduğu değişkenlerin korelasyon analizlerine bakıldığında her üç yöntemde de aynı değişkenler arasında anlamlı ilişkiler bulunmuştur. Korelasyon katsayısı değerleri incelendiğinde nitel ve nicel değişkenlerin korelasyon katsayıları üç yöntemde de birbirlerine yakın bulunmuş iken tamamı nitel değişken olanlar arasındaki korelasyon katsayıları üç yöntemde de birbirine eşit çıkmıştır.

5. BAĞIMLI DEĞİŞKENİ NİTEL OLAN İSTATİSTİK YÖNTEMLER

Basit ve çoklu doğrusal regresyon yöntemlerinin kullanılacağı verilerin sürekli veya kesikli sayısal değer olması gerekir. Bağımlı ve bağımsız değişkenlerin her ikisi de normal dağılım göstermelidir. Bu gibi koşullar sağlanmadığında basit ya da çoklu doğrusal regresyon analizi kullanılamaz. Bağımsız değişkenlerden biri veya bir kaçı nitel ise kukla değişkeni yardımıyla kurulan yapay değişkenli modeller tercih edilir. Bağımlı değişken nitel ise bağımsız değişkenlerin normal dağılım varsayımını sağlamaması durumunda lojistik regresyon, bağımsız değişkenlerin normal dağılım varsayımını sağlaması durumunda ise diskriminant analizi tercih edilir. Bu bölümde bağımlı ve bağımsız değişkenlerin normalliği sağlama veya sağlamama durumlarına göre tercih edilen regresyon modelleri karşılaştırılmıştır. Veri olarak üçüncü bölümde normal dağılım varsayımları test edilen veri setleri kullanılmıştır.

5.1. Yapay Değişkenli Modeller

Regresyon analizinde kullanılan değişkenler her zaman istikrarlı bir seyir göstermeyebilir. Araştırma dönemleri içinde araştırılan konuyu ciddi olarak etkileyen keskin uçlar olabilir. Ekonomik durumun savaş öncesi ile sonrası arasında veya kandaki bir maddenin hastalık öncesi ile sonrası arasında aşırı farklar olabilir. Bu gibi farkların etkisini göstermek için 0 ve 1 değerlerini alan yapay değişkenler kullanılır. Yapay değişkeni; kukla, gölge veya yardımcı değişken olarak isimlendiren kaynaklar da vardır. Verileri kategorilere ayırdığı için "kategorik veri" ifadesi de kullanılmaktadır.

Yapay değişkenler genellikle bağımsız değişken olarak kullanılmaktadır ancak bağımlı değişken olduğu durumlar da mevcuttur. Varyans ve kovaryans analiz modelleri, bağımsız değişkenin yapay olduğu durumlarda başlıca yöntemler olarak kullanılırlar (Tarı, 2008; Akkaya ve Pazarlıoğlu, 1998).

Yapay değişkenin bağımlı değişken olarak yer aldığı çalışmalarda sıklıkla kullanılan yöntemler probit ve lojistik regresyon modelleridir. Özellikle gerekli hesaplamaları yapabilen bilgisayar programlarından sonra yapay bağımlı değişkenli modeller için lojistik regresyon analizi ve probit modeller en sık kullanılan yöntemler olmuştur (Gujarati, 2005).

5.2. Probit Model

İki kategorili bağımlı değişkeni incelemek için uygun seçilmiş birikimli dağılım fonksiyonlarını kullanmak gerekmektedir. Örneğin; lojistik modelde lojistik dağılım fonksiyonu kullanılmaktadır. Bazı durumlarda normal birikimli dağılım fonksiyonları da kullanılabilir. Bu fonksiyonu kullanan yapay bağımlı değişkenli modellerden biri de fayda teorisi üzerine kurulmuş probit modeldir. Normit model olarak da bilinen probit model herhangi bir konuda i. bireyin kararının gözlenemeyen bir fayda endeksine bağlı olduğu varsayımı ile hareket eder. Bağımsız değişkene bağlı olarak belirlenen endeks ne kadar büyük olursa istenilen olayın gerçekleşme ihtimali o kadar yüksek olur (Akın, 2002).

Endeks;

$$I_i = \beta_1 + \beta_2 X_i \tag{5.1}$$

olarak ifade edilir. Denklemde X_i i. gözlemin değeridir. Doğrusal olasılık modelinde $Y_i=1$ ile istenilen olayın gerçekleşmesi ifade edilir. $Y_i = 0$ ile olayın gerçekleşmemesi ifade edilmektedir. Probit modelde ise istenilen olayın gerçekleşmesinin I_i kritik değeri belirlenir. Herhangi bir i. gözlem için gözlemin kritik endeks değeri I_i^* ise ve olayın gerçekleşmesi için "1" değeri atanmışsa eşitsizlik şu şekilde yazılabilir:

$I_i^* \leq I_i ise\ Y = 1$ olayı gerçekleşecek.

$I_i^* \geq I_i ise\ Y = 1$ olayı gerçekleşmeyecek.

Doğrudan gözlenemeyen I_i^* normal dağılımlı kabul edilirse standart birikimli normal dağılım fonksiyonundan aşağıdaki gibi hesaplanabilir.

$$P_i = P(Y = 1) = P(I_i^* \leq I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{\frac{-t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1+\beta_2 X_i} e^{\frac{-t^2}{2}} dt \tag{5.2}$$

Eşitlikte t standart normal dağılımdır. P_i olayın gerçekleşme ihtimali -∞ dan I_i ye kadar olan standart normal eğri altında kalan alan hesabı ile bulunur. Yukarıda yazılan eşitlik I_i'nin bir fonksiyonu olarak aşağıdaki gibi yazılabilir.

$$F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1+\beta_2 X_i} e^{\frac{-t^2}{2}} dt \tag{5.3}$$

Bu fonksiyonun doğrusal olmadığı görülüyor. Bu fonksiyonel eşitliğin tersi alınıp doğrusallaştırılarak endeks elde edilir.

$$F^{-1}\big(F(I_i)\big) = \beta_1 + \beta_2 X_i \tag{5.4}$$

Probit model parametre tahminlerinin karmaşıklığı ve yorumlamasındaki zorluktan dolayı logit model kadar tercih edilmez (Kutlar, 2007).

Literatürde probit model kullanılan çalışmalar incelendiğinde; Altıntaş ve Duru (2007) probit ve lojistik model ile marka bağımlılığı üzerinde çalışma yapmışlardır. Cebeci (2012) krizleri incelemede nitel tercih modelleri üzerinde çalışmış ve Türkiye için bir uygulama yapmıştır. Kırcı Çevik ve Korkmaz (2014) Türkiye'de yaşam doyumu ile iş doyumu arasındaki ilişkiyi sıralı probit model ile analiz etmiştir. Akkaya ve Kantar da (2018) finansal tahminlerle ilgili çalışmalarında probit ve logit modeli kullanmışlardır. Erdugan ve Yörübulut (2018) yatan hastaların hastaneyi tekrar tercih etme durumlarını probit regresyon analizi ile araştırmışlardır.

5.3. Lojistik Regresyon Analizi

Regresyon analizi, doğrusal olma ve doğrusal olmama durumuna göre ikiye ayrılmaktadır. Doğrusallık, "parametrelere" veya "değişkenlere" göre doğrusallık olmak üzere iki şekilde incelenir. Değişkenlerde doğrusallık durumu, değişkenlerin tamamının birinci dereceden kuvvet ile yazılmasını zorunlu kılar. Değişkenler iki ve daha fazla kuvvet ya da köklü ifade ile yazılmamalıdır. Değişkenler birbiri ile çarpım veya bölüm halinde bulunmamalıdır. Benzer şekilde parametrelere göre doğrusallıkta da katsayılar için aynı şartlar aranır. Fakat bazı modeller doğrusal görünmese bile bazı dönüşümlerle doğrusal hale getirilebilmektedir. Bu sebeple "özünde doğrusal olan" ve "özünde doğrusal olmayan" şeklinde bir kavram kullanılabilir. Log-doğrusal, yarı logaritmik ve ters modeller doğrusala dönüştürülebilen yaygın modellerdir (Gujarati, 2005).

Probit ve lojistik model birbirine oldukça benzerdir. Aralarındaki temel fark kullanmış oldukları fonksiyonlardır. Uygulamalarda da birbirlerine oldukça yakın sonuçlar verirler. Ancak matematiksel uygulaması ve yorumlama kolaylığından dolayı lojistik model daha çok tercih edilmektedir. Paket programların yaygınlaşması ile probit

modelin bahsedilen zorluğu aşılmaktadır. Kullanılacak olan paket programa ve araştırmacının tercihine göre iki model arasında tercih yapılabilir (Gujarati, 2005).

Lojistik regresyon modeli logaritmik dönüşümler sonucu doğrusal hale gelen, bağımlı değişkeni kategorik, yani yapay değişken olan bir modeldir. Lojistik regresyon bağımlı ve bağımsız değişkenler arasında logit bir ilişki olduğunu varsayar; dolayısıyla lojistik regresyon doğrusal olmayan modeller üretebilir.

Lojistik regresyon analizinin tercih edilme nedenlerini şöyle özetleyebiliriz:

1. Bağımlı değişken kategoriktir. Bağımsız değişkenlerin sürekli ya da süreksiz olmasına yönelik kısıtlama getirmemektedir.
2. Lojistik modelde model doğrusal hale getirilerek model kurulumu kolaylaştırılmaktadır.
3. Lojistik regresyon analizleri yapabilen paket programların yaygındır. (SPSS, SAS vb.)
4. Bağımsız değişkenlerin olasılık fonksiyonlarının dağılımıyla ilgili bir şart bulunmaz.
5. Lojistik regresyonda negatif olasılıkla karşılaşma sorunu olmaz.
6. Lojistik regresyon analizinde, bağımsız değişken ile bağımlı değişken arasındaki ilişkinin doğrusal olması zorunlu değildir. Bu değişkenler arasındaki ilişki üstel veya polinom ilişkisi de olabilir.

1972 yılında Anderson bireylerden tespit edilen özelliklerden çoğunun veya hepsinin kalitatif yapıda olması durumunda grupları birbirinden ayırmada veya sınıflandırmada kullanılabilecek en uygun diskriminant metodu üzerinde durmuştur. Bu gibi durumlarda lojistik ayırıcı katsayıların en uygun sonuçlar verdiğini açıklamıştır. Aynı yıl Atkison ikili cevap değişkeni içeren doğrusal lojistik modelin uygunluğu için bir test geliştirmiş ve bu testi eş karşılaştırmalarında kullanılan Bradley-Terry modeline uygulamıştır. Ayrıca bu yöntemin küçük örneklerdeki sonuçlarını görebilmek amacıyla sayısal bir uygulama yapmıştır. Prentice (1976), geriye dönük planlanmış çalışmalarda, bağımlı değişkenin hastalık durumunu gösterdiği (var veya yok) ve bağımsız değişkenin iki seviyeli kategorik bir değişken olarak belirlendiği durumlarda lojistik regresyon modelinin kullanımı üzerinde durmuştur. Bu amaçla model kurma aşamalarını, hesaplama adımlarının ve sonuçların yorumlanmalarını, vaka-kontrol denemesinden elde edilen bir veri seti üzerinde uygulamalı olarak göstermiştir. Diğer taraftan Pregibon

(1981), yaptığı çalışmada lojistik regresyon modeli ve benzeri modellerde, sapan ve etkili olan gözlemlerin tespitinde kullanılabilecek tanı istatistikleri üzerinde durmuştur. 1982 yılında, Kleinbaum ve ark. lojistik regresyon analizinin epidemiyolojik çalışmalarda kullanım amaçlarını ele alarak; bu analize ilişkin temel açıklamalar, basit doğrusal lojistik regresyon modeli, bu modelin özel durumları ve modeldeki katsayıların şartlı olabilirlik tahminleri üzerinde durmuşlar ve epidemiyoloji dalında elde edilen bir veri setinde bu tekniğin uygulama adımlarını ve sonuçlarının yorumlamasını göstermişlerdir. Begg ve Gray (1984), ikiden fazla seviye içeren cevap değişkeni için kurulan lojistik regresyon modellerinin, basit lojistik regresyon modelleri kullanılarak analiz edilebilmesi olanakları üzerinde durmuşlardır. Bu modellerin kullanımı sonucunda elde edilen katsayıların asimptotik dağılımlarını elde etmişler ve ikiden çok seviye içeren cevap değişkeninin yer aldığı modeldeki katsayıların en çok olabilirlik tahmin edicileri ile karşılaştırmışlardır. Sonuçta ise basit lojistik regresyon analizinin, katsayı tahminlerinde uygun olduğunu vurgulamışlardır. Diğer taraftan Carroll ve ark. 1984 yılında, yaptıkları çalışmada iki seviye içeren cevap değişkenine sahip lojistik regresyon modellerinde, değişkenlerin hatalı ölçüldüğü durumlardaki tahmin yöntemleri üzerinde durmuşlardır. Normal ölçüm hatası içeren değişkenlere ait katsayıların tahmininde yapısal en çok tahmin metotlarının, ölçüm hatalarının büyük olduğu durumlarda ise şartlı en çok olabilirlik tahmin metotlarının özelliklerini araştırmışlardır. Johnson (1985), lojistik regresyon analizinde etkili olabilecek gözlemler üzerinde durmuştur. Bu amaçla önerdiği ölçüleri diğer çalışmalarda kullanılan olabilirlik uzaklığı ve sapma gibi ölçülerle karşılaştırmıştır. Ayrıca teorik açıklamaları da bir veri seti üzerinde uygulamalı olarak göstermiştir. Bonney (1987), yaptığı çalışmada iki seviye içeren bir cevap değişkeninin olabilirliğini, açıklayıcı değişken içeren ve içermeyen modellerdeki şartlı olasılıkların çarpımı olarak ifade etmiş ve bu modelleri regressive logistik modeller olarak adlandırmıştır. Katsayıların tahmininde en çok olabilirlik yöntemini kullanmış, yapılan teorik açıklamaları üç farklı veri seti üzerinde uygulamalı olarak denemiş ve verilerin analizinde istatistik paket programlarından yararlanmıştır. Robert ve ark. (1987), anket denemelerinde kümeleme veya tabakalaşmalardan dolayı verilerin analizinde kullanılan standart χ^2 istatistiği veya olabilirlik oran istatistiğinin, birinci tip hata değerlerinin etkilendiğini vurgulamışlardır. Bu sorunu gidermek amacıyla anket denemesinin önemine göre bir takım düzeltmeler yaparak lojistik

regresyon analizinin kullanılabileceğini göstermişlerdir. Cox ve Snell (1989), iki seviyeli cevap değişkenine ait özellikler üzerinde durarak, bu tip bir değişkene ait kurulacak modelin lojistik analizini açıklamışlardır. Ayrıca, çapraz tablolarda ve çok sayıda açıklayıcı değişken içeren modellerde lojistik regresyon analizinin kullanımı üzerinde de durmuşlardır. Hosmer ve ark. (1989), çok sayıda açıklayıcı değişken içeren lojistik regresyon modelleri için en uygun değişken kombinasyonu üzerinde durmuşlardır. Yapılan çalışma sonucunda herhangi bir en iyi alt seti veren doğrusal regresyon programı kullanılarak lojistik regresyon için en iyi alt seti belirlemede elde edilen performansın, hataları normal dağılım göstermeyen modeller kullanılarak belirlenen en iyi alt setin performansına eş değer olduğunu göstermişlerdir. 1989'da Hosmer ve Lemeshow, basit lojistik regresyon modelinin tanıtımı, katsayıların önemi ve tahmin testi, çoklu lojistik regresyon modeli, modelde yer alan katsayıların yorumlanması, uygun modeli kurma aşamaları, uyum iyiliği testleri, vaka-kontrol çalışmalarında lojistik regresyon modelinin kullanımı, cevap değişkeninin ikiden çok seviye içerdiği durumlar için kurulan lojistik regresyon modelinin analizi ve hayatta kalma verileri için lojistik regresyon analizinin uygulanışı üzerine temel teorik ve uygulamalı açıklamalarda bulunmuşlardır. Dobson (1990), genelleştirilmiş doğrusal modellerin bir üyesi olan lojistik regresyon modeli üzerinde temel açıklamalarda bulunmuştur. Lojistik modelde yer alan cevap değişkeninin sadece iki seviye içerdiği durumlardaki katsayıların en çok olabilirlik ve en küçük kareler tahminleri, olasılık dağılımları, uyum iyiliği kriterleri ve doz-tepki modelleri üzerinde durmuştur. Diğer taraftan 1990 yılında Başarır, diskriminant ve lojistik regresyon analizlerinde; verilerin yapısındaki grup sayısı bilinmekte ise buna göre bir ayrımsama modeli kurmuştur. Kurulan bu model yardımı ile veri kümesine yeni alınan gözlemlerin gruplara atanması da yapılabilmektedir. O'Neil ve Barry (1995), yaptıkları çalışmada iki seviyeli cevap değişkeninin görülme olasılığını gösteren yani başarılı veya genel olarak var şeklinde ifade edilen seviyesinin nadir görüldüğü durumlarda, lojistik regresyon modelleri ve katsayı tahmin yöntemleri üzerinde durmuşlardır. Denemelerde özellikle trafik kazaları verilerini dikkate alıp ölüm sebebi olabilecek yaş, cinsiyet, aracın hız limiti gibi bağımsız değişkenlerin etkilerini araştırmışlardır. Bircan (2004), ikili sonuç değişkeni ile hem sürekli hem de kesikli değişkenlerden oluşan bağımsız değişkenler kümesi arasındaki ilişkiyi tanımlayabilen lojistik regresyon analizinin incelenmesi ile ilgili;

lojistik regresyon analizine bir uygulama göstermek amacıyla, çocuklarda doğum ağırlığını etkileyen risk faktörleri üzerinde çalışmıştır. Coşkun ve ark. (2004), lojistik regresyon analizi ile diş hekimliği konusunda bir uygulama yapmıştır. Kaya ve ark. (2007), yaptıkları çalışmada tıp fakültesi ve sağlık yüksekokulu öğrencilerinde depresif belirtilerin yaygınlığını, stres ile başa çıkma tarzlarını ve etkileyen faktörleri lojistik regresyon analizi ile incelemişlerdir. Girginer ve Cankuş (2008), tramvay yolcu memnuniyetinin lojistik regresyon analiziyle ölçülmesi üzerinde çalışmışlardır. Aktaş ve ark. (2009), bağımlı değişkenin iki düzeyli olması durumunda demografik, davranış ve risk faktörüyle ilgili tahmin çalışmalarında oldukça sık kullanılan lojistik regresyon analizini incelemişler ve parametre tahminine ilişkin yöntemi ayrıntılı bir şekilde inceleyerek, katsayıların yorumu için odds oranını açıklamışlardır. Kitiş ve ark. (2009), yaptıkları çalışmada yirmi yaş ve üzeri kadınlarda metabolik sendrom sıklığını ve bunu etkileyen faktörleri; tek yönlü varyans analizi, korelasyon analizi ve lojistik regresyon analizi yöntemleriyle incelemişlerdir. Çokluk (2010), lojistik regresyonun kavramlarını açıklayarak eğitim bilimleri üzerine bir uygulamasını yapmıştır. Diğer taraftan Yıldız (2011), serebral palsili çocuklarda konvülsiyonu etkileyen risk faktörlerini lojistik regresyon analizi ile incelemiştir.

Bağımlı değişkenin yapay olduğu modellerin temeli olan doğrusal olasılık modelinin sorunlarını aşmak için bazı dönüşümlerle doğrusal hale gelen fonksiyonlara ihtiyaç vardır. Lojistik dağılım fonksiyonunu incelemek için iki kategorili bağımlı değişken ele alınabilir.

$$P_i = E(Y = 1/X_i) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_i)}} \tag{5.5}$$

$Z_i = \beta_0 + \beta_1 X_i$ dönüşümü yapılırsa

$P_i = \frac{1}{1+e^{-Z_i}}$ elde edilir. (5.6)

Bu eşitlik "Lojistik Dağılım Fonksiyonu" olarak bilinir. Eşitlikte Z_i, $-\infty$ ile $+\infty$ arasında değer alırken P_i'nin de 0 ile 1 arasında değerler alacağı açıktır. Z_i'nin dolayısıyla da X_i'nin P_i ile ilişkisinin doğrusal olmadığı da eşitlikte görülmektedir. Bu iki özelliğin sağlanması ile doğrusal olasılık modelinin eksiklikleri giderilmiş olmaktadır (Gujarati, 2005).

Belli bir X^* değerinde istenilen olayın gerçekleşme ihtimali lojistik dağılım fonksiyonunda yerine yazılarak bulunur. Lojistik modeldeki β_1 katsayısı ise X'deki 1 birimlik artışın L'de yapacağı artışı yani istenilen olayın gerçekleşme lehine fark oranındaki artışı gösterir (Akkaya ve Pazarlıoğlu, 1998).

$$P_i = \frac{1}{1+e^{-Z_i}} \tag{5.7}$$

Eşitliğini tekrar ele alıp doğrusal hale getirilmesi incelenebilir. P_i, kategorik değişkene ait istenilen olayın gerçekleşmesi ihtimalini ifade eder. O halde olayın gerçekleşmeme ihtimali ise $(1- P_i)$'dir.

$$(1 - P_i) = \frac{1}{1+e^{Z_i}} \tag{5.8}$$

olarak bulunur. Olayın gerçekleşmesi olasılığının gerçekleşmeme olasılığına oranı ise;

$$\frac{P_i}{1-P_i} = \frac{1+e^{Z_i}}{1+e^{-Z_i}} = e^{Z_i} \text{ 'dir.} \tag{5.9}$$

Bu oran olayın gerçekleşmesinin gerçekleşmeme olasılığına fark oranı veya bahis oranı olarak adlandırılmaktadır. Bu eşitlikte her iki tarafın doğal logaritması alınarak lojistik model elde edilir.

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = \ln e^{Z_i} \tag{5.10}$$

$Z_i = \beta_0 + \beta_1 X_i$ değeri yerine yazılırsa,

$$L_i = \ln(e^{\beta_0+\beta_1 X_i}) = \beta_0 + \beta_1 X_i \text{ olarak bulunur.} \tag{5.11}$$

Bu dönüşüm sonrası L, hem kat sayılara göre hem de değişkenlere göre doğrusal hale gelmiş oldu. "L" logit olarak adlandırıldığı için logit model ismi de buradan gelmektedir (Gujarati, 2005).

Lojistik regresyon analizinde modelin uygunluğu "model ki-kare" testi ile, her bağımsız değişken için anlamlılık testi ise Wald istatistiği ile test edilir.

Odds, oddsratio ve logit, lojistik regresyon analizinin önemli temel kavramlarıdır.

Odds: Bir olayın görülme ihtimalinin, görülmeme ihtimaline oranıdır.

Oddsratio (OR): İki odds'nin birbirine oranıdır. Bu oran iki değişken arasındaki ilişkiyinin özetidir.

Logit: Odds ratio'nun doğal logaritmasının alınmasıdır. Odds ratio asimetrik olduğundan doğal logaritması alınıp simetrik hale getirilir. Logit, doğrusal regresyon analizindeki β katsayısıdır.

Lojistik regresyon analizi genellikle bağımlı değişkenin iki kategorili olduğu durumlarda kullanılır. Ancak bağımlı değişkenin ikiden fazla kategorili nitelik olarak belirlendiği aşağıdaki Tablo 5.1'deki durumlarda da lojistik regresyon analizi uygulanabilir.

Tablo 5.1. Kategori Sayısına Göre Lojistik Regresyon Örnekleri

Bağımlı Değişken	**Örnek**
2 Kategorili (Binominal)	öldü-yaşıyor, başarılı-başarısız
2'den fazla kategorili sırasız (Multinominal)	işsiz-emekli-çalışan, sayısal-sözel-eşit ağırlık
2'den fazla kategorili sıralı (Ordinal)	düşük-orta-yüksek, etkisiz-etkili-çok etkili

Lojistik regresyon analizi ile çalışmak için sağlanması gereken varsayımlar olmasa da araştırmalarda aşağıda belirtilen genel istatistik kurallarına uyulması faydalı olacaktır.

• Uygun olan tüm bağımsız değişkenler modele dahil edilmelidir.

Bazı değişkenlerin modele dahil edilmemesi hata teriminin büyümesine ve modelin yetersizliğine neden olabilir.

• Uygun olmayan tüm bağımsız değişkenler dışlanmalıdır.

Nedensel olarak uygun olmayan değişkenlerin modele dahil edilmesi, modeli karmaşık hale getirebilir. Modelin yorumlanmasını zorlaştırabilir. Bu değişkenlerin bağımlı değişken üzerinde etkisi varmış gibi yanlış izlenime neden olabilir.

• Aynı gözlem üzerinde bir kez veri alınmalı, tekrarlayan ölçümler olmamalıdır.

• Bağımsız değişkenlerde ölçüm hatası küçük olmalıdır.

Ölçüm hataları küçük olmalı, eksik veri olmamalıdır. Hatalar, katsayıların tahmininde yanlılığa ve modelin yetersizliğine neden olur.

• Bağımsız değişkenler arasında çoklu bağlantı olmamalıdır.

Bağımsız değişkenler birbirleriyle ilişkili olmamalıdır.

•Aşırı değerler olmamalıdır.

Aşırı değerler sonucu önemli derecede etkileyebilir.

•Örneklem büyüklüğü yeterli olmalıdır.

Az sayıda gözlem içeren örneklemde tahmin edilen değerlerin güvenilirliği azalır.

• Beklenen ve gözlenen varyanslar arasındaki fark küçük olmalıdır.

Bağımlı değişkenin beklenen varyansı ile gözlenen varyansı arasındaki büyük bir fark modelin yetersizliğini gösterir. Bu, örneklemin rastgele seçilmesinden veya araştırılan konuda akademik eksiklik gibi bir sorundan kaynaklanmaktadır.

5.4. Diskriminant Analizi

Diskriminant analizi değişkenlerden elde edilen fonksiyonlar ile gruplar arasında maksimum ayrım yapmayı hedefleyen yöntemdir. Bu fonksiyonlar yardımı ile yeni gelen bir birimin hangi gruba dahil olması gerektiğine karar verilir (Toktay, 2017).

Diskriminant analizi, çeşitli değişkenlere göre iki veya daha fazla örnek grubu arasındaki farklılıkları belirlemeye çalışan bir istatistiksel tekniktir. Birimlerin gruplanması için matematiksel eşitlikler kullanılır. Diskriminant fonksiyonu denilen bu eşitlikler, birbirine en çok benzeyen grupları belirlemeye çalışmaktadır. Bunun için ortak özellikler belirlenerek benzer gözlemler aynı gruba toplanır. Grupları ayıran karakteristikler diskriminant değişkenleri olarak adlandırılmaktadır. Yani diskriminant analizi, diskriminant değişkenleri yardımıyla iki veya daha fazla grubun farklılıklarının belirlenmesi işlemidir. Diskriminant fonksiyonları, değişkenlerin doğrusal bileşenlerinden oluşur. Farklılığa neden olan değişkenlerin hangileri olduğunu belirlemeye çalışır. Diskriminant analizi ile hangi gruba ait olduğu bilinmeyen bir birimin ait olduğu grup en az hata ile belirlenmektedir. Diskriminant analizi, grupların farklılaşmasına en fazla etki eden değişkenleri belirleyerek grupların farklılaşmasında etkili değişkenlerin saptanmasını da sağlar. Gerçek grup üyelikleri ile analiz sonucunda yapılan sınıflandırmanın karşılaştırılması sonucunda elde edilen diskriminant fonksiyonun yeterli ve güvenilir olup olmadığını belirlemeye olanak sağlar. Diskriminant analizinin çalışma mantığı çok değişkenli varyans analizinde olduğu gibidir. Grupları ortalamalarına göre ortak ortalamadan farklı olmalarını sağlayacak bir ayırıcı kriter geliştirmeyi amaçlar.

Diskriminant analizinin uygulanabilmesi için bazı koşulların sağlanması gereklidir.

- Veriler çok değişkenli normal dağılım göstermelidir.
- Değişkenlerin varyans ve kovaryansları homojen olmalıdır. Değişkenler ortak kovaryans matrisine sahip, çok değişkenli ana kütleden çekilmiş örnekler olmalıdır.
- Değişkenlerin ortalamaları ve varyansları arasında bir korelasyon bulunmamalıdır.
- Değişkenler arasında çoklu bağlantı olmamalıdır.
- Değişkenler grupların birbirinden ayrılmasında etkin olmayacak kadar gereksiz olmamalı. Grupların birbirinden ayrılmasını sağlayacak kadar doğru ve gerekli değişkenleri içermelidir.

Diskriminant fonksiyonu katsayıları hesaplanırken farklı yöntemler kullanılabilir. Diskriminant analizi bazı araştırmacılar ve kaynaklarda kullanılan yöntemlere göre farklı adlarla anılmaktadır. Fisher'in doğrusal diskriminant analizi, Bayes diskriminant analizi, Kernel tabanlı kümeleme ile diskriminant analizi, Laplacian doğrusal diskriminant analizi en büyük benzerlik diskriminant analizi, Kuadratik diskriminant analizi gibi isimler kullanılmaktadır (Burmaoğlu vd., 2009).

Diskriminant analizinde bağımlı değişken iki ya da daha fazla kategorili değişkendir. Bağımsız değişkenler ise sürekli ya da kesikli nicel verilerdir. Diskriminant analizinde bağımlı değişken doğru seçilmelidir. Gruplar birbiriyle ilişkisiz, tam bağımsız olmalı ve eksiksiz bir şekilde tanımlanmalıdır. Yani her gözlem sadece bir gruba yerleşebilmelidir. Gruplara ayırmada bir sınırlama getirilmemektedir. Ancak genelde bağımlı grup sayısı iki, üç ya da dört olmaktadır. Birbiriyle tam bağımsız, iyi ayrılmış grupların daha fazlası olması halinde de yöntemin sorunsuz çalışacağı düşünülmektedir. Bağımlı değişken yani gruplar hasta-sağlam, sayısal-sözel-eşit ağırlık şeklinde nominal veri olabileceği gibi çok az, az, orta, çok şeklinde sıralı nitelik veri de olabilir.

Diskriminant analizinin kullanılacağı araştırmalarda bağımlı değişken yukarıdaki gibi belirlendikten sonra bağımsız değişkenlerin seçimine geçilir. Diskriminant analizi ile gruplar arasındaki farklılığı maksimum yapan bağımsız değişkenleri bulmak ve bağımsız değişkenlerin farklılıklarından yararlanarak da grup üyelikleri tahmin etmek amaçlanır. O halde diskriminant analizinin başarısı bağımsız değişkenlerin seçimi ile

yakından ilgilidir. Doğru seçim için grup üyeliğine ayırmada başarılı olacak, gruplar arasındaki ayırt ediciliği en yüksek bağımsız değişkenler seçilmelidir.

Değişken seçimleri öncelikle araştırmacının bilgi birikimine, konuya olan hakimiyetine bağlıdır. Bununla birlikte maliyet, bilgiye ulaşma kolaylığı gibi uygulamada kolaylık sağlayacak özelliklerde dikkate alınmalıdır. Bağımsız değişkenlerin seçiminde göz ardı edilemeyecek diğer bir husus ise konu ile ilgisi olduğu düşünülen bilgi ve önseziden faydalanıp daha önceki çalışmalardan da yararlanmaktır.

Diskriminant analizi çalışmalarında örneklem büyüklüğü için gözlem ve bağımsız değişken sayısına bakılır. Öncelikli olarak sınıflandırılmış gruplardan en az gözleme sahip olanın bağımsız değişkenlerin sayısından daha fazla olması istenir. Örneklem büyüklüğünün, bağımsız değişken sayısının en az 4 katı olması gerektiği de söylenmektedir. Her grubun gözlem sayısının en az 20 olması gerektiği görüşünde olanlar da vardır. Her bir grup için eşit sayıda gözlem yapılması zorunluluk değildir. Sınıflandırma aşamasında gözlem sayısı fazla olana daha fazla atama yapılacağından tavsiye edilmemektedir (Alpar, 2017). Fakat yukarıda yazılanlarla beraber normallik şartlarını sağlayacak kadar gözlem olması gerektiği de göz ardı edilmemelidir. Bağımsız değişken sayısının çok az olması halinde gözlem sayısı da az olacaktır. Oysa merkezi limit teoremine göre gözlem sayısı arttıkça dağılım normale yaklaşacaktır. Bu genel istatistik kuralları da unutulmamalıdır. Aksi halde alternatif yöntemler kullanılmalıdır. Bu hususlara dikkat edilerek yapılacak bir çalışmada gözlem sayısı fazla olana daha fazla atama yapılması da ciddi bir sorun yaratmayacaktır. Çünkü analiz sonunda kıyaslamalar genellikle oranlara göre yapılarak yorumlanmaktadır.

Diskriminant analizi MANOVA'nın tersi gibi düşünülebilir. MANOVA ile var olan gruplar arasında fark var mı diye araştırılırken, diskriminant analizinde benzerlik var mı diye araştırma yapılmaktadır. MANOVA'da gruplar bağımsız değişken, gözlem değerleri bağımlı değişkendir. Diskriminant analizinde ise gruplar bağımlı değişken, gözlem değerleri bağımsız değişkenlerdir.

Lojistik regresyon analizi ile diskriminant analizi arasında da benzerlikler bulunmaktadır. Her ikisi de bağımlı değişkenin kategorik veri olması durumunda kullanılır. Fakat diskriminant analizi bağımsız değişkenler için bazı kısıtlamalar getirmektedir. Diskriminant analizinde bağımsız değişkenlerin sürekli veya kesikli nicel veri tipinde olması gerekirken, lojistik regresyon analizinde böyle bir şart olmayıp

nitelik veriler de bağımsız değişken olabilir. Diskriminant analizi varyans-kovaryans matrisi homojenliği, çok değişkenli normallik gibi varsayımlar gerektirirken, lojistik regresyon analizinde bu gibi ön şartlar bulunmamaktadır. Varsayımların karşılanmadığı veya veri tipinin diskriminant analizine uymadığı durumlarda lojistik regresyon analizi tercih edilebilir.

Diskriminant analizinde ikili gruplandırma yapılıyorsa "iki gruplu diskriminant analizi" ikiden fazla grup varsa "çoklu diskriminant analizi" denilir. Diskriminant analizine başlarken çalışma konusuna ilişkin bağımsız değişkenler ile grupların ayırt edilip edilemeyeceği yani gruplar arasında farkın var olup olmadığına karar verilir. Fark olması kararı ile birlikte diskriminant fonksiyonları da oluşmuş olur. Ayırt edicilikte hangi değişkenlerin rolünün en fazla olduğu da belirlenmiş olmaktadır. Değişkenler gruplandırmayı yapamayacaksa, çalışmaya devam edilmez.

Diskriminant analizi varsayımlarından çok değişkenli normallik ve varyans-kovaryans homojenliği sağlanmıyorsa dönüşümlerden yararlanılabilir. Verilerin normal dağılımı sağlaması ancak varyans-kovaryans matrislerinin homojen olmaması durumunda kuadratik diskriminant analizi önerilmektedir. Diskriminant analizi, varyans-kovaryans matrislerinin benzerliğine karşı duyarlıdır. Yetersiz örneklem büyüklüğü ve benzer olmayan varyans-kovaryans matrisi olması durumunda sınıflandırma işleminin olumsuz etkileneceği söylenmektedir. Bağımsız değişkenlerden bazıları arasında yakın ilişki olması çoklu bağlantı sorununa neden olur. Böyle değişkenlerin diskriminant analizinin amacı olan ayırt ediciliğe katkısı oldukça az olacaktır. Varsayımlara göre gruplar içerisindeki bağımsız değişken çiftlerinin tümü arasında doğrusal bir ilişki vardır. Bu varsayımın sağlanmaması testin gücünü azaltır. Dönüşüm yapılarak devam edilecek analiz sonucunda ise dönüşüm uygulanan değişkenlerin yorumlanması zorlaşabilmektedir. Bu tür değişkenlerin mümkün ise değiştirilmesi değilse analizden çıkartılması da düşünülebilir. İstatistik analizlerin tümünde olduğu gibi yine aşırı değerlere de dikkat etmek gerekir (Alpar, 2017).

Matematiksel eşitlik olarak gösterilen diskriminant fonksiyonları birbirine en çok benzeyen grupları belirlemeye çalışmaktadır. Diskriminant analizi birimlere ait özelliklerden yararlanarak ait oldukları grupları belirlemeye veya var olan grupları birbirinden ayırabilecek en iyi fonksiyonu bulmak için kullanılan tekniklerinden biridir (Toktay, 2017)

Atakan ve Karabulut (2003), derinliğe dayalı diskriminasyon çalışmışlardır. Selim ve Sarıbay (2003), Yabancı dil çalışan öğrencilerin beklentileri üzerine çalışma yapıp, diskriminant analizini kullanmışlardır. Oğuzlar (2006), hanehalkı tipi üzerindeki kır-kent ayrımı çalışmasında diskriminant analizi çalışmıştır. Tektaş (2014), diskriminant analizi ile lojistik ragresyon analizinin sınıflandırma başarısı üzerine bir çalışma yapmıştır. Çelik vd. (2015), diskriminant analizini bıldırcınlar üzerinde yapmış oldukları çalışmada kullanmışlardır.

5.5. Lojistik Regresyon Analizi, Probit Analizi ve Diskriminant Analizi Karşılaştırması

Bağımlı ve bağımsız değişken ayrımının yapıldığı çalışmalarda kullanılacak yönteme karar verebilmek için değişkenlerin türüne bakılır. Bağımlı değişkenin yapay değişken olduğu durumlarda lojistik regresyon analizi, probit analizi ve diskriminant analizi kullanılabilir. Bu yöntemler, bağımlı değişkenin türüne göre benzer iken bağımsız değişkelerin türüne göre farklılıklar gösterirler. Diskriminant analizinde bağımsız değişkenlerin parametrik olma şartlarını taşıması gerekirken diğer iki yöntem böyle bir varsayım gerektirmemektedir. Bu durumda ilk olarak diskriminant analizinde bağımsız değişkenlerin nicel veri olması gerekmektedir. Diğer iki yöntemde bağımsız değişkenin yapay değişken olması durumunda da analize devam edilebilir. Bağımsız değişkenler sayısal veri olduğu halde parametrik olma şartlarında ciddi sapmalar olması durumunda da lojistik regresyon analizi ve probit analizi önerilmektedir.

Araştırmalarda kullanılan veriler parametrik olma şartlarının tamamını aynı anda sağlayamayabilirler. Normal dağılıma uygun bir veri varyans homojenliği varsayımını karşılamayabilir. Bu ikisini sağladığı halde değişkenler arasında bağlantı olmama şartını sağlamayabilirler. Varsayımların tümünü sağlayan verilere ulaşmak genellikle zordur. Farklı varsayımları sağlama durumlarına göre uygulanacak yöntemlerin başarıları da farklı olabilmektedir. Bu çalışmanın ilerleyen bölümlerinde varsayımlar ve bağımlı değişkendeki kategori sayısına göre farklılık gösteren veriler üzerinde analizler yapılarak karşılaştırılacaktır.

5.5.1. Tek ve Çok Değişkenli Normal Dağılıma Uygun; Varyans ve Kovaryans Homojenliği Olmayan Verilerle Analiz

Üçüncü bölümde DATA-1 ile verilerek normal dağılım varsayımı test edilen veriler kullanılarak analizler yapılacaktır. Bu verilere göre bir mağazada çalışanlar üç farklı bölüme sınıflandırılmaktadır. Bu sınıflar bağımlı değişken olan müşteri temsilcisi, muhasebe ve makinacıdan oluşan üç kategorili değişkendir. Bu sınıflandırmayı yapmak için çalışanların tecrübe, giyim tarzı ve bilgisayar bilme düzeyleri puanlandırılarak değerlendirme yapılmaktadır. Bu üç değişken de analizin bağımsız değişkenlerini oluşturan nicel verilerdir. Yani DATA-1; tek ve çok değişkenli normal dağılıma uygun ve üç kategorili bağımlı değişkene sahip verilerdir.

Diskriminant analizinin uygulanabilmesi için gerekli önemli varsayımlardan biri de varyans, kovaryans homojenliğidir. Varyans homojenliği ve çoklu bağlantı da test edilerek diskriminant analizinin sonuçları incelenecektir.

Değişkenler arasında çoklu bağlantıyı kontrol etmek için Tablo 5.2'ye bakılabilir.

Tablo 5.2. DATA-1 Korelasyon Matrisi

		Tecrübe	**Giyim Tarzı**	**Bilgisayar Düzeyi**
Korelasyon	Tecrübe	1	0,071	-0,077
	Giyim Tarzı	0,071	1	0,015
	Bilgisayar Düzeyi	-0,077	0,015	1

Tablodan değişkenler arasında çoklu bağlantının olmadığı söylenebilir. 0,70 den fazla olan korelasyonlar için çoklu bağlantı sorunu olduğu söylenebilir.

Varyans ve kovaryans homojenliği için Tablo 5.3 ve Tablo 5.4 incelenebilir.

Tablo 5.3. DATA-1 Varyans Homojenliği Testi

	Levene İstatistik	**sd1**	**sd2**	**Sig.**
Tecrübe	6,236	2	242	0,002
Giyim Tarzı	0,332	2	242	0,718
Bilgisayar Düzeyi	11,493	2	242	0,000

Varyansların iki değişken için homojen olmadığı sadece bir değişkenin varyans homojenliği şartını taşıdığı görülür.

Tablo 5.4. DATA-1 Box's M Testi

Box's M		37,433
F	**Approx.**	3,062
	sd1	12
	sd2	239037,539
	Sig.	0,000

Box's M tablosundan kovaryans homojenliği şartının sağlanmadığı görülebilir.

5.5.1.1. Diskriminant Analizi: Tek ve Çok Değişkenli Normal Dağılıma Uygun; Varyans ve Kovaryans Homojenliği Olmayan Veriler

Diskriminant fonksiyonlarına ait özdeğer istatistikleri Tablo 5.5'de verilmiştir.

Tablo 5.5. DATA-1 Özdeğerler

Fonksiyon	Özdeğer	Varyans Yüzdesi	Birikimli Yüzde	Kanonik Korelasyon
1	0,474	95,1	95,1	0,567
2	0,024	4,9	100,0	0,154

Üç kategorili bağımlı değişken olduğundan Tablo 5.5'de iki fonksiyon elde edilmiştir. Büyük özdeğer istatistiği bağımlı değişkendeki varyansın fonksiyon tarafından iyi açıklanacağını göstermektedir. Kesin olmamakla birlikte 0,40'dan büyük değerler iyi kabul edilmektedir. Bu veriler için elde edilen özdeğerlerin çok iyi olmadığı tablodan görülebilir. Benzer olarak kanonik korelasyon değerleri de düşük çıkmıştır. Bu değerlerin karesi bağımlı değişkendeki varyansın açıklama yüzdesini göstermektedir. $0,567^2 = 0,32$ ve $0,154^2 = 0,02$ değerleri düşüktür. Kısaca Tablo 5.5'e göre şimdiden iyi bir sınıflama yüzdesi olmayacağı söylenebilir.

Tablo 5.6. DATA-1 Wilk's Lambda İstatistiği

Test fonksiyonları	Wilks' Lambda	Ki-kare	sd	Sig.
1'den 2'ye	0,662	99,264	6	0,000
2	0,976	5,813	2	0,055

Tablo 5.6'da gruplar arasındaki farkla açıklanamayacak olan oranı gösteren Wilks' Lambda değerleri %66 ve %98 olarak hesaplanmış ve birinci fonksiyon anlamlı bulunmuştur.

Tablo 5.7. DATA-1 Standartlaştırılmış Kanonik Diskriminant Fonksiyon Katsayıları

	Fonksiyon	
	1	**2**
Tecrübe	0,046	0,632
Giyim Tarzı	0,991	-0,149
Bilgisayar Düzeyi	0,085	0,822

Tablo 5.7'ye göre her iki fonksiyonda da değişkenlerin tümü modele dahil edilmiştir. Birinci fonksiyon için en iyi ayırt edici olan değişkenin giyim tarzı, ikinci fonksiyon için bilgisayar düzeyi olduğu görülüyor.

Tablo 5.8. DATA-1 Yapı Matrisi

	Fonksiyon	
	1	**2**
Giyim Tarzı	0,996	-0,092
Bilgisayar Düzeyi	0,096	0,771
Tecrübe	0,110	0,558

Her bir değişkenin diskriminant fonksiyonu ile olan korelasyonunu gösteren Tablo 5.8'den bağımsız değişkenlerin önemine bakılabilir. Birinci fonksiyon için en iyi ayırt edici olan değişkenin giyim tarzı, ikinci fonksiyon için bilgisayar düzeyi olduğu bir önceki tabloda da belirtilmişti.

Tablo 5.9. DATA-1 Standartlaştırılmamış Kanonik Diskriminant Fonksiyonu Katsayıları

	Fonksiyon	
	1	**2**
Tecrübe	0,009	0,123
Giyim Tarzı	0,212	-0,032
Bilgisayar Düzeyi	0,022	0,213
Sabit	-4,688	-2,775

Tablo 5.9'daki katsayılar ile diskriminant fonksiyonları oluşturulur. Buna göre:

Z_1 = - 4,688 + 0,009x(tecrübe) + 0,212x(giyim tarzı) + 0,022x(bilgisayar düzeyi)

Z_2 = - 2,775 + 0,123x(tecrübe) - 0,032x(giyim tarzı) + 0,213x(bilgisayar düzeyi)

Tablo 5.10. DATA-1 Diskriminant Analizi Sınıflandırma Yüzdeleri

Meslek			Tahmin edilen grup üyelikleri			Toplam
			Müşteri temsilcisi	Muhasebe	Makinacı	
Orjinal	Sayı	Müşteri temsilcisi	51	23	11	85
		Muhasebe	32	43	18	93
		Makinacı	5	20	42	67
	%	Müşteri temsilcisi	60,0	27,1	12,9	100,0
		Muhasebe	34,4	46,2	19,4	100,0
		Makinacı	7,5	29,9	62,7	100,0
Sınıflandırma başarısı: %55,5						

Tablo 5.10 yapılan analizin başarısını gösteren bir tablodur. Farklı veri tipleri ve farklı yöntemlerin karşılaştırıldığı bu çalışma için de önemlidir. Analiz sonucunda bağımsız değişkenlerin grupları ayırmadaki başarısının %55,5 olduğu görülüyor.

5.5.1.2. Lojistik Regresyon Analizi: Tek ve Çok Değişkenli Normal Dağılıma Uygun; Varyans ve Kovaryans Homojenliği Olmayan Veriler

Üçüncü bölümde normalliği incelenen DATA-1; normal dağılıma uygun, varyans homojenliği olmadığı belirlenen üç kategorili bağımlı değişkene sahip bir lojistik regresyon modeli olarak göz önüne alınabilir. Bir önceki başlıkta aynı veriler için incelenen diskriminant analizinin sonuçları ile karşılaştırmak için hiçbir varsayıma ihtiyaç duymayan lojistik regresyon analizi sonuçları elde edilmiştir.

Tablo 5.11. DATA-1 Model Uyum Bilgisi

Model	Model uyum kriteri	Olabilirlik oran testi		
	-2 Log Likelihood	Ki-Kare	sd	Sig.
Sadece kesişim	532,487			
Final	433,116	99,372	6	0,000

Tablo 5.11'de bulunan anlamlılık değeri model uyumunun iyi olacağını göstermektedir.

Tablo 5.12. DATA-1 Pseudo R^2

Cox and Snell	0,333
Nagelkerke	0,376

Tablo 5.12'ye bakarak bağımlı değişken ile bağımsız değişkenler arasında Cox and Snell'e göre yaklaşık %33, Nagelkerke'ye göre %37 ilişki olduğu söylenebilir.

Tablo 5.13. DATA-1 Lojistik Regresyon Analizi Parametre Tahminleri

Meslek[a]		B	Std. Hata	Wald	sd	Sig.	Exp (B)	Exp(B) için %95 güven aralığı	
								Alt Sınır	Üst Sınır
Müşteri temsilcisi	Kesişim	-7,418	1,153	41,427	1	0,000			
	Tecrübe	0,013	0,040	0,104	1	0,748	1,013	0,937	1,095
	Giyim Tarzı	0,366	0,049	56,741	1	0,000	1,442	1,311	1,587
	Bilgisayar Düzeyi	0,025	0,052	0,235	1	0,628	1,026	0,926	1,136
Muhasebe	Kesişim	-5,164	0,986	27,455	1	0,000			
	Tecrübe	0,050	0,036	1,941	1	0,164	1,051	0,980	1,127
	Giyim Tarzı	0,224	0,042	28,788	1	0,000	1,250	1,152	1,357
	Bilgisayar Düzeyi	0,089	0,046	3,679	1	0,055	1,093	0,998	1,196
a. Referans kategorisi: Makinacı.									

Tablo 5.13'de lojistik regresyon analizinin de diskriminant analizinde olduğu gibi anlamlı bulunan değişkeni giyim tarzı olarak belirlediği görülebilir.

Tablo 5.14. DATA-1 Lojistik Regresyon Analizi Sınıflandırma Yüzdeleri

Gözlem Değerleri	Tahmin değerleri			
	Müşteri temsilcisi	Muhasebe	Makinacı	Doğru sınıflama yüzdesi
Müşteri temsilcisi	50	24	11	%58,8
Muhasebe	32	43	18	%46,2
Makinacı	5	20	42	%62,7
Toplam sınıflama yüzdesi	%35,5	%35,5	%29,0	%55,1

Tablo 5.14'de lojistik regresyon analizinin doğru sınıflandırma oranının % 55,1 olduğu görülür. Aynı veriler için %55,5 doğru sınıflama oranı yapan diskriminant analizi yöntemine oldukça yakın sınıflandırma yaptığı söylenebilir.

5.5.2. Tek Değişkenli Normalliği Sağlamayıp, Çoklu Normalliği Sağlayan; Varyans ve Kovaryans Homojenliği Olan Verilerle Analiz

Üçüncü bölümde bağımlı değişkeni üç kategorili olarak verilen DATA-2 verisinin kategori sayısı ikiye düşürülüp DATA-6 verileri elde edilmiştir. Bu verilerden faydalanarak diskriminant, lojistik regresyon ve probit analizleri karşılaştırılacaktır. Bağımlı değişkendeki kategori sayısına göre yöntemlerin sınıflandırma başarıları da gözlemlenmiş olacaktır. Ayrıca adımsal yöntemler ile enter yöntemleri arasında fark olup olmadığı da kontrol edilecektir. DATA-2 verileri tek değişkenli normalliğe yakın görülmeyip çoklu normal dağılıma uyan verilerdi. Bu verilerin puan değerleri değiştirilmediğinden normalliğe ilişkin bir değişiklik olmayacaktır.

Tablo 5.15. DATA-6 Korelasyon Matrisi

		Tecrübe	**Giyim Tarzı**	**Bilgisayar Düzeyi**
Korelasyon	Tecrübe	1,000	-0,003	0,064
	Giyim Tarzı	-0,003	1,000	-0,232
	Bilgisayar Düzeyi	0,064	-0,232	1,000

Tablo 5.15'den değişkenler arasında çoklu bağlantının olmadığı söylenebilir. 0,70'den fazla olan korelasyonlar için çoklu bağlantı sorunu olduğu söylenebilir.

Tablo 5.16. DATA-6 Varyans Homojenliği Testi

	Levene İstatistik	**df1**	**df2**	**Sig.**
Tecrübe	2,335	1	243	0,128
Giyim Tarzı	3,754	1	243	0,054
Bilgisayar Düzeyi	3,024	1	243	0,083

Tablo 5.16 ile DATA-6 verilerinin varyans homojenliği kontrol edilmiş varyans homojenliği şartını sağladığı görülmüştür.

Tablo 5.17. DATA-6 Box's M testi

Box's M		8,058
F	**Approx.**	1,325
	df1	6
	df2	427749,081
	Sig.	0,242

Tablo 5.17 Box's M tablosundan kovaryans homojenliği şartının sağlandığı da görülebilir. DATA-6 verileri tek değişkenli normalliğe yakın görülmeyip çoklu normal dağılıma uyan, varyans ve kovaryans homojenliği olan değişkenleri arasında önemli bir korelasyon bulunmayan verilerdir.

5.5.2.1. Diskriminant Analizi: Tek Değişkenli Normalliği Sağlamayıp, Çoklu normalliği Sağlayan; Varyans ve Kovaryans Homojenliği Olan Veriler

Diskriminant fonksiyonuna ait özdeğer ve kanonik korelasyon değerleri Tablo 5.18'de verilmiştir.

Tablo 5.18. DATA-6 Özdeğerler

Fonksiyon	Özdeğer	Varyans Yüzdesi	Birikimli Yüzde	Kanonik Korelasyon
1	0,293	100,0	100,0	0,476

Büyük özdeğer istatistiği bağımlı değişkendeki varyansın fonksiyon tarafından iyi açıklanacağını göstermektedir. Kesin olmamakla birlikte 0,40'dan büyük değerler iyi kabul edilmektedir. Bu veriler için elde edilen özdeğerin çok iyi olmadığı tablodan görülebilir. Benzer olarak kanonik korelasyon değeri de düşük çıkmıştır. Bu değerin karesi bağımlı değişkendeki varyansın açıklama yüzdesini göstermektedir. $0,476^2$ = 0,23 değeri düşüktür.

Tablo 5.19. DATA-6 Wilk's Lambda İstatistiği

Test fonksiyonları	Wilks' Lambda	Ki-kare	sd	Sig.
1	0,774	61,979	3	0,000

Tablo 5.19'da gruplar arasındaki farkla açıklanamayacak olan oranı gösteren Wilks' Lambda değeri %77 olarak hesaplanmış ve fonksiyon anlamlı bulunmuştur.

Diskriminant analizi, kurulacak olan fonksiyona alınacak değişkenleri aşağıdaki Tablo 5.20 ile belirlemiştir.

Tablo 5.20. DATA-6 Standartlaştırılmış Kanonik Diskriminant Fonksiyon Katsayıları

	Birinci fonksiyon
Tecrübe	0,962
Giyim Tarzı	-0,285
Bilgisayar Düzeyi	-0,037

Tablo 5.20'ye göre diskriminant fonksiyonunda değişkenlerin tümü modele dahil edilmiştir. Fonksiyon için en iyi ayırt edici olan değişkenin tecrübe olduğu görülüyor.

Tablo 5.21. DATA-6 Yapı Matrisi

	Birinci fonksiyon
Tecrübe	0,960
Giyim Tarzı	-0,279
Bilgisayar Düzeyi	0,090

Her bir değişkenin diskriminant fonksiyonu ile olan korelasyonunu gösteren Tablo 5.21'den bağımsız değişkenlerin önemine bakılabilir. Fonksiyon için en iyi ayırt edici olan değişkenin tecrübe olduğu bir önceki tabloda da belirtilmişti.

Kurulacak olan model fonksiyonu için ise Tablo 5.22'daki katsayılar alınacaktır.

Tablo 5.22. DATA-6 Standartlaştırılmamış Kanonik Diskriminant Fonksiyonu Katsayıları

	Birinci fonksiyon
Tecrübe	0,224
Giyim Tarzı	-0,053
Bilgisayar Düzeyi	-0,010
Sabit	-2,308

Tablo 5.22'deki katsayılar ile diskriminant fonksiyonu oluşturulur.

Z = -2,308 + 0,224 x (tecrübe) – 0,053 x (giyim tarzı) – 0,010 x (bilgisayar düzeyi)

Tablo 5.23 ile diskriminant analizinin DATA-6 verilerini sınıflandırmadaki başarısı verilmiştir.

Tablo 5.23. DATA-6 Diskriminant Analizi Sınıflandırma Yüzdeleri

Meslek			**Tahmin edilen grup üyelikleri**		**Toplam**
			Muhasebe	**Müşteri temsilcisi**	
Orjinal	**Sayı**	Muhasebe	91	32	123
		Müşteri temsilcisi	40	82	122
	%	Muhasebe	74,0	26,0	100,0
		Müşteri temsilcisi	32,8	67,2	100,0
Sınıflandırma başarısı: %70,6					

Tabloya göre diskriminant analizinin DATA-6 verilerini kullanarak yaptığı doğru sınıflandırma oranı %70,6'dır.

Aşağıda Tablo 5.24, Tablo 5.25, Tablo 5.26'da DATA-6 verilerine diskriminant analizinin adımsal olarak uygulanması durumundaki sonuçları verilmiştir.

Tablo 5.24. DATA-6 Adımsal Yöntemde Standartlaştırılmış Kanonik Diskriminant Fonksiyon Katsayıları

	Birinci fonksiyon
Tecrübe	0,960
Giyim Tarzı	-0,277

Tablo 5.25. DATA-6 Adımsal Yöntemde Standartlaştırılmamış Kanonik Diskriminant Fonksiyonu Katsayıları

	Birinci fonksiyon
Tecrübe	0,224
Giyim Tarzı	-0,051
Sabit	-2,440

Tablo 5.26. DATA-6 Adımsal Yöntem Diskriminant Analizi Sınıflandırma Yüzdeleri

Meslek			**Tahmin edilen grup üyelikleri**		**Toplam**
			Muhasebe	**Müşteri temsilcisi**	
Orjinal	**Sayı**	Muhasebe	91	32	123
		Müşteri temsilcisi	41	81	122
	%	Muhasebe	74,0	26,0	100,0
		Müşteri temsilcisi	33,6	66,4	100,0
Sınıflandırma başarısı %70,2					

Adımsal yöntem ile yapılan analizde model kurmak için üç bağımsız değişkenden ikisi; tecrübe ve giyim tarzı modele dahil edilerek bilgisayar düzeyi değişkeni modele alınmamıştır. Enter yöntemiyle yapılan analizde sınıflandırma yüzdesini %70,6 iken adımsal metotta %70,2 doğru sınıflandırma yapılmıştır.

5.5.2.2. Lojistik Regresyon Analizi: Tek Değişkenli Normalliği Sağlamayıp, Çoklu Normalliği Sağlayan; Varyans ve Kovaryans Homojenliği Olan Veriler

DATA-6 verileri kullanılarak incelenen diskriminant analizinin sonuçları ile karşılaştırmak için aynı verilere, hiçbir varsayıma ihtiyaç duymayan lojistik regresyon analizi uygulanmıştır.

Tablo 5.27. DATA-6 Model Özeti

Adım	-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
1	275,678	0,230	0,306

Tablo 5.27'ye bakarak bağımlı değişken ile bağımsız değişkenler arasında Cox and Snell'e göre yaklaşık %23, Nagelkerke'ye göre %30 ilişki olduğu söylenebilir. Lojistik regresyon analizi ile kurulacak model için aşağıdaki Tablo 5.28'den yararlanılır.

Tablo 5.28. DATA-6 Lojistik Regresyon Analizi Denklem Değişkenleri

		B	S.E.	Wald	df	Sig.	Exp(B)
Adım 1[a]	Tecrübe	-0,253	0,040	40,779	1	0,000	0,776
	Giyim Tarzı	0,060	0,028	4,490	1	0,034	1,062
	Bilgisayar Düzeyi	0,014	0,040	0,119	1	0,730	1,014
	Sabit	2,596	0,973	7,119	1	0,008	13,417
a. Variable(s) entered on step 1: Tecrübe, Giyim Tarzı, Bilgisayar Düzeyi.							

Lojistik regresyon analizinin anlamlı bulduğu değişkenler tecrübe ve sabit terimdir. Değişkenlerin modeldeki önemlerini de diskriminant analizinde olduğu gibi sırasıyla tecrübe, giyim tarzı, bilgisayar düzeyi olarak belirlemiştir.

$$L = 2{,}596 - 0{,}253 \text{ x (tecrübe)} + 0{,}060 \text{ x (giyim tarzı)} + 0{,}014 \text{ x (bilgisayar düzeyi)}$$

Tablo 5.29'de lojistik regresyon analizinin sınıflandırma başarısı verilmiştir. Tabloya göre lojistik regresyon analizinin DATA-6 verilerini kullanarak yaptığı doğru sınıflandırma oranı %70,6'dır.

Tablo 5.29. DATA-6 Lojistik Regresyon Analizi Sınıflandırma Yüzdeleri

Gözlenen değerler			Tahmin değerleri		
			Meslek		Doğru sınıflama yüzdesi
			Muhasebe	Müşteri temsilcisi	
Adım 1	Meslek	Muhasebe	90	33	73,2
		Müşteri temsilcisi	39	83	68,0
	Toplam sınıflama yüzdesi				70,6

Aşağıda Tablo 5.30 ve Tablo 5.31'de DATA-6 verilerine lojistik regresyon analizinin adımsal olarak uygulanması durumundaki sonuçları verilmiştir.

Tablo 5.30. DATA-6 Lojistik Regresyon Analizi Adımsal Yöntemde Denklemdeki Değişkenler

		B	S.E.	Wald	df	Sig.	Exp(B)
Adım 1[a]	Tecrübe	-0,253	0,040	40,779	1	0,000	0,776
	Giyim Tarzı	0,060	0,028	4,490	1	0,034	1,062
	Bilgisayar Düzeyi	0,014	0,040	0,119	1	0,730	1,014
	Sabit	2,596	0,973	7,119	1	0,008	13,417
Adım 2[a]	Tecrübe	-0,252	0,039	40,766	1	0,000	0,777
	Giyim Tarzı	0,057	0,027	4,420	1	0,036	1,059
	Sabit	2,776	0,825	11,318	1	0,001	16,055
a. Variable(s) entered on step 1: Tecrübe, Giyim Tarzı, Bilgisayar Düzeyi.							

Tablo 5.31. DATA-6 Adımsal Yöntemde Lojistik Regresyon Analizi Sınıflandırma Yüzdeleri

Gözlenen değerler			Tahmin değerleri		
			Meslek		Doğru sınıflama yüzdesi
			Muhasebe	Müşteri temsilcisi	
Adım 1	Meslek	Muhasebe	90	33	73,2
		Müşteri temsilcisi	39	83	68,0
	Toplam Sınıflama Yüzdesi				70,6
Adım 2	Meslek	Muhasebe	88	35	71,5
		Müşteri temsilcisi	38	84	68,9
	Toplam sınıflama yüzdesi				70,2

Adımsal yöntem ile yapılan analizlerde hem diskriminant analizi hem de lojistik regresyon analizi model kurmak için aynı değişkenleri seçmiştir. Üç bağımsız değişkenden ikisi tecrübe ve giyim tarzı, modele dahil edilerek bilgisayar düzeyi değişkeni modele alınmamıştır. Yine her iki yöntemin sınıflandırma yüzdeleri de aynı olmuştur. Enter yöntemiyle yapılan analizlerde her iki yöntem de sınıflandırma yüzdesini %70,6 bulmuşken adımsal metotta ise her iki yöntem de %70,2 doğru sınıflandırma yapmıştır. Sınıflandırma açısından enter yöntemi ile adımsal yöntemler arasında önemli bir fark bulunmazken değişken sayısının azaltılarak model kurulmak istenmesi durumunda adımsal yöntemler tercih edilebilir.

Başlık 5.5.1'de incelenmiş olan üç kategorili bağımlı değişken olması durumunda doğru sınıflandırma oranının lojistik regresyon analizinde %55,1 diskriminant analizinde %55,5 olarak birbirlerine oldukça yakın değerler aldığı söylenebilir. Bu başlıkta incelenen iki kategorili bağımlı değişkenle analizde ise her iki yöntem de sınıflandırma yüzdesini %70,6 bulmuştur. Bu çalışmada kullanılan veriler için, iki yöntemin sınıflandırma başarılarının bağımlı değişkendeki kategori sayısına göre farklılık göstermediği söylenebilir.

5.5.2.3. Probit Analizi: Tek Değişkenli Normalliği Sağlamayıp, Çoklu Normalliği Sağlayan; Varyans ve Kovaryans Homojenliği Olan Veriler

Diskriminant ve lojistik regresyon analizi ile DATA-6 verileri için model kurarken değişkenlerin etki sıralamalarının aynı olduğu görülmüştür. Probit analizinde değişkenlerin etki sıralamasının nasıl olduğunu görmek için Tablo 5.32'den yararlanılır.

Tablo 5.32. DATA-6 Probit Analizi Parametre Tahminleri

Parametre		**Tahmin**	**Std. Hata**	**Z**	**Sig.**	**%95 Güven aralığı**	
						Alt sınır	**Üst sınır**
PROBIT	**Tecrübe**	-0,152	0,053	-2,880	0,004	-0,255	-0,048
	Giyim Tarzı	0,038	0,024	1,571	0,116	-0,009	0,084
	Bilgisayar Düzeyi	0,008	0,024	0,318	0,750	-0,040	0,056
	Kesişim	1,528	0,573	2,666	0,008	0,955	2,102

Probit regresyonun anlamlı bulduğu değişkenler lojistik regresyon analizi ile aynı olup, tecrübe ve sabit terimdir. Değişkenlerin modeldeki önem dereceleri her üç yöntemde de aynı bulunmuş olup, sırasıyla tecrübe, giyim tarzı, bilgisayar düzeyi olarak belirlenmiştir.

Y = 1,528 - 0,152 x (tecrübe) + 0,038 x (giyim tarzı) + 0,008 x (bilgisayar düzeyi)

5.5.3. Tek Değişkenli Normalliği Sağlamayıp, Çoklu Normalliği Sağlayan; Varyansları Homojen, Kovaryansları Homojen Olmayan Verilerle Analiz

Bu başlıkta kullanılacak veriler üçüncü bölümde DATA-2 olarak verilip normalliği test edilen verilerdir. Verilerin tek değişkenli normal dağılıma uymadığı çok değişkenli normalliği sağladığı ilgili kısımda belirtilen verilere ait analiz karşılaştırmaları yapılacaktır.

Tablo 5.33'den değişkenler arasında çoklu bağlantının olmadığı söylenebilir. 0,70'den fazla olan korelasyonlar için çoklu bağlantı sorunu olduğu söylenebilir.

Tablo 5.33. DATA-2 Korelasyon Matrisi

		Tecrübe	**Giyim Tarzı**	**Bilgisayar Düzeyi**
Korelasyon	Tecrübe	1,000	0,079	0,019
	Giyim Tarzı	0,079	1,000	0,060
	Bilgisayar Düzeyi	0,019	0,060	1,000

Tablo 5.34 ile DATA-6 verilerinin varyans homojenliği kontrol edilmiştir.

Tablo 5.34. DATA-2 Varyans Homojenliği Testi

	Levene İstatistik	**df1**	**df2**	**Sig.**
Tecrübe	2,219	2	242	0,111
Giyim Tarzı	1,364	2	242	0,258
Bilgisayar Düzeyi	0,756	2	242	0,470

Test sonuçlarına göre varyans homojenliği tüm değişkenler için sağlanmıştır.

Tablo 5.35'e göre kovaryansların homojen olmadığı söylenebilir.

Tablo 5.35. DATA-2 Box's M Testi

Box's M		26,119
F	**Approx.**	2,137
	df1	12
	df2	239037,539
	Sig.	0,012

DATA-2 verileri tek değişkenli normalliğe uymayan, çoklu normal dağılıma uyan, varyansları homojen, kovaryansları homojen olmayan verilerdir.

5.5.3.1. Diskriminant Analizi: Tek Değişkenli Normalliği Sağlamayıp, Çoklu Normalliği Sağlayan; Varyansları Homojen, Kovaryansları Homojen Olmayan Veriler

Tablo 5.36 ile özdeğerler ve kanonik korelasyon değerleri verilmiştir.

Tablo 5.36. DATA-2 Özdeğerler

Fonksiyon	Özdeğer	Varyans Yüzdesi	Birikimli Yüzde	Kanonik Korelasyon
1	1,052	76,8	76,8	0,716
2	0,317	23,2	100,0	0,491

Büyük özdeğer istatistiği bağımlı değişkendeki varyansın fonksiyon tarafından iyi açıklanacağını göstermektedir. Kesin olmamakla birlikte 0,40'dan büyük değerler iyi kabul edilmektedir. Bu veriler için özdeğerlerin iyi olduğu söylenebilir. Benzer olarak kanonik korelasyon değerleri de iyi çıkmıştır. Bu değerlerin karesi bağımlı değişkendeki varyansın açıklama yüzdesini göstermektedir. $0,716^2 = 0,51$; $0,491^2=0,24$

Tablo 5.37 düşük Wilks' Lamda değeri açıklanamayacak kısmın düşük olacağını göstermektedir.

Tablo 5.37. DATA-2 Wilk's Lambda İstatistiği

Test fonksiyonları	Wilks' Lambda	Ki-kare	sd	Sig.
1'den 2'ye	0,370	239,630	6	0,000
2	0,759	66,432	2	0,000

Gruplar arasındaki farkla açıklanamayacak olan oranı gösteren Wilks' Lambda değerleri yaklaşık %37 ve %75 olarak hesaplanmış fonksiyonlar anlamlı bulunmuştur.

Tablo 5.38 ile DATA-2 verilerinin diskriminant analizi sonucundaki doğru sınıflandırma yüzdeleri verilmiştir.

Tablo 5.38. DATA-2 Diskriminant Analizi Sınıflandırma Yüzdeleri

Meslek			**Tahmin edilen grup üyelikleri**			**Toplam**
			Müşteri temsilcisi	**Muhasebe**	**Makinacı**	
Orjinal	**Sayı**	Müşteri temsilcisi	67	14	4	85
		Muhasebe	16	67	10	93
		Makinacı	3	14	50	67
	%	Müşteri temsilcisi	78,8	16,5	4,7	100,0
		Muhasebe	17,2	72,0	10,8	100,0
		Makinacı	4,5	20,9	74,6	100,0
Sınıflandırma başarısı %75,1						

Doğru sınıflandırma oranı %75,1 olmuştur.

5.5.3.2. Lojistik Regresyon Analizi: Tek Değişkenli Normalliği Sağlamayıp, Çoklu Normalliği Sağlayan; Varyansları Homojen, Kovaryansları Homojen Olmayan Veriler

DATA-2 verileri için incelenen diskriminant analizinin sonuçları ile karşılaştırmak üzere hiçbir varsayıma ihtiyaç duymayan lojistik regresyon analizi sonuçları elde edilmiştir. Model bilgisine ait Tablo 5.39'un anlamlı bulunması model uyumunun iyi olacağını göstermektedir.

Tablo 5.39. DATA-2 Model Uyum Bilgisi

Model	**Model Uyum Kriteri**	**Olabilirlik Oran Testi**		
	-2 Log Likelihood	**Ki-Kare**	**sd**	**Sig.**
Sadece Kesişim	532,487			
Final	285,738	246,749	6	0,000

Tablo 5.40 ile bağımlı değişken ile bağımsız değişkenler arasındaki ilişki verilmiştir. Bu ilişki Cox and Snell'e göre yaklaşık %64, Nagelkerke'ye göre %72 olarak tespit edilmiştir.

Tablo 5.40. DATA-2 Pseudo R^2

Cox and Snell	0,635
Nagelkerke	0,716

Tablo 5.41 ile lojistik regresyon modeli için parametre tahminleri verilmiştir.

Tablo 5.41. DATA-2 Lojistik Regresyon Analizi Parametre Tahminleri

Meslek[a]		**B**	**Std. Hata**	**Wald**	**sd**	**Sig.**	**Exp (B)**	**Exp(B) için %95 güven aralığı**	
								Alt Sınır	**Üst Sınır**
Müşteri temsilcisi	Kesişim	-2,314	1,571	2,168	1	0,141			
	Tecrübe	-0,255	0,070	13,404	1	0,000	0,775	0,676	0,888
	Giyim Tarzı	0,554	0,074	56,533	1	0,000	1,740	1,506	2,011
	Bilgisayar Düzeyi	-0,454	0,086	27,937	1	0,000	0,635	0,537	0,752
Muhasebe	Kesişim	-5,506	1,468	14,064	1	0,000			
	Tecrübe	0,209	0,056	14,029	1	0,000	1,232	1,105	1,375
	Giyim Tarzı	0,325	0,060	29,031	1	0,000	1,384	1,230	1,558
	Bilgisayar Düzeyi	-0,322	0,071	20,497	1	0,000	0,725	0,631	0,833
Referans kategorisi: Makinacı									

Tablo 5.42'de lojistik regresyon analizinin doğru sınıflandırma oranı verilmiştir.

Tablo 5.42. DATA-2 Lojistik Regresyon Analizi Sınıflandırma Yüzdeleri

Gözlenen değerler	**Tahmin değerleri**			
	Müşteri temsilcisi	**Muhasebe**	**Makinacı**	**Doğru sınıflama yüzdesi**
Müşteri temsilcisi	67	14	4	%78,8
Muhasebe	15	67	11	%72,0
Makinacı	3	14	50	%74,6
Toplam sınıflama yüzdesi	%34,7	%38,8	%26,5	%75,1

DATA-2 verileri için lojistik regresyon da doğru sınıflandırma oranını %75.1 olarak belirlemiştir. Bu veriler için lojistik regresyon analizi ile diskriminant analizi sınıflandırma oranını aynı bulmuştur.

5.5.4. Tek Değişkenli Normalliğe Uygun, Çoklu Normalliği Sağlamayan; Varyansları Homojen, Kovaryansları Homojen Olmayan Verilerle Analiz

Bu başlıkta, üçüncü bölümde DATA-3 ile verilerek normal dağılım varsayımı test edilen veriler kullanılarak analizler yapılacaktır. Bu verilere göre bir spor okulunda öğrencilerin yatkın oldukları spor dalını belirlemek için açık atlama, smaç, krospas, kondüsyon ve sıçrama değişkenlerine ilişkin puanlamalar yapılmakta, aldıkları puanlara göre öğrenciler jimnastik veya voleybola yönlendirilmektedirler. Diskriminant analizinin uygulanabilmesi için gerekli önemli varsayımlardan biri de varyans, kovaryans homojenliğidir. Varyans homojenliği ve çoklu bağlantı da test edilerek diskriminant analizinin sonuçları incelenecektir.

Değişkenler arasında çoklu bağlantıyı kontrol etmek için Tablo 5.43 incelenebilir.

Tablo 5.43. DATA-3 Korelasyon Matrisi

		Açık Atlama	**Smaç**	**Krospas**	**Kondüsyon**	**Sıçrama**
Korelasyon	Açık Atlama	1,000	-0,995	-0,854	0,024	0,886
	Smaç	-0,995	1,000	0,855	-0,022	-0,885
	Krospas	-0,854	0,855	1,000	-0,030	-0,844
	Kondüsyon	0,024	-0,022	-0,030	1,000	0,024
	Sıçrama	0,886	-0,885	-0,844	0,024	1,000

Tabloya göre değişkenler arasında çoklu bağlantının olduğu söylenebilir. 0,70 den fazla olan korelasyonlar için çoklu bağlantı sorunu olduğu söylenebilir.

Varyans ve kovaryans homojenliği için Tablo 5.44 ve Tablo 5.45 incelenebilir.

Tablo 5.44. DATA-3 Varyans Homojenliği

	Levene İstatistik	**df1**	**df2**	**Sig.**
Açık Atlama	3,789	1	483	0,052
Smaç	3,588	1	483	0,059
Krospas	1,594	1	483	0,207
Kondüsyon	1,652	1	483	0,199
Sıçrama	3,479	1	483	0,063

Tablodan değişkenlerin varyans homojenliği şartını taşıdığı görülür.

Tablo 5.45. DATA-3 Box's M Testi

Box's M		358,535
F	**Approx.**	23,632
	df1	15
	df2	816858,057
	Sig.	0,000

Box's M tablosundan kovaryans homojenliği şartının sağlanmadığı görülebilir. DATA-3 tek değişkenli normalliğe uygun, çoklu normalliği sağlamayan; varyansları homojen, kovaryansları homojen olmayan; değişkenleri arasında çoklu bağlantı olan verilerdir.

5.5.4.1. Diskriminant Analizi: Tek Değişkenli Normalliğe Uygun, Çoklu Normalliği Sağlamayan; Varyansları Homojen, Kovaryansları Homojen Olmayan veriler

Tablo 5.46'de diskriminant fonksiyonuna ait özdeğer ve kanonik korelasyon değerleri verilmiştir.

Tablo 5.46. DATA-3 Özdeğerler

Fonksiyon	Özdeğer	Varyans Yüzdesi	Birikimli Yüzde	Kanonik Korelasyon
1	1,964[a]	100,0	100,0	0,814

Yaklaşık 0,66 olan kanonik korelasyon değerinin karesi ($0,814^2$= 0,66) modelin bağımlı değişkendeki varyansın %66'sını açıklayabildiğini göstermektedir. Büyük özdeğer istatistiği de (1,964) model uyumunun iyiliğini göstermektedir.

Tablo 5.47'de düşük Wilks' Lamda değeri açıklanamayacak kısmın düşük olacağını gösteriyor.

Tablo 5.47. DATA-3 Wilk's Lambda İstatistiği

Test fonksiyonları	Wilks' Lambda	Ki-kare	sd	Sig.
1	0,337	522,095	5	0,000

Yaklaşık %34 olan Wilks'Lambda değeri açıklanamayan kısmın oranını göstermekte olup, anlamlı signification değeriyse elde edilecek diskriminant fonksiyonunun anlamlı olduğunu göstermektedir.

Diskriminant analizi kurulacak olan fonksiyona alınacak değişkenleri aşağıdaki Tablo 5.48 ile belirlemiştir.

Tablo 5.48. DATA-3 Standartlaştırılmış Kanonik Diskriminant Fonksiyon Katsayıları

	Birinci fonksiyon
Açık Atlama	1,171
Smaç	-0,214
Krospas	0,027
Kondüsyon	-0,021
Sıçrama	-0,431

Tabloya göre grupları ayırt etmede her bir bağımsız değişken önemli bulunmuştur. Her bir değişkenin diskriminant fonksiyonu ile olan korelasyonunu gösteren değerlere Tablo 5.49'den bakılabilir.

Tablo 5.49. DATA-3 Yapı Matrisi

	Birinci fonksiyon
Açık Atlama	0,979
Smaç	-0,975
Krospas	-0,792
Sıçrama	0,773
Kondüsyon	0,000

Yukarıdaki değerler bağımsız değişkenlerin fonksiyondaki önemini gösteren verilerdir. Kurulacak olan model fonksiyonu için ise Tablo 5.50 'daki katsayılar alınacaktır.

Tablo 5.50. DATA-3 Standartlaştırılmamış Kanonik Diskriminant Fonksiyonu Katsayıları

	Birinci fonksiyon
Açık Atlama	0,390
Smaç	-0,072
Krospas	0,008
Kondüsyon	-0,005
Sıçrama	-0,093
Sabit	-2,346

Diskriminant fonksiyonunu oluşturmada yukarıdaki değerler bağımsız değişkenlerin katsayıları olarak kullanılırlar.

Z = -2,346 + 0,390x(açık atlama) – 0,072x(smaç) + 0,008x(krospas) – 0,005x(kondüsyon) – 0,093x(sıçrama)

Tablo 5.51'de DATA-3 verilerinin diskriminant analizi sonucundaki doğru sınıflanma yüzdeleri verilmiştir.

Tablo 5.51. DATA-3 Diskriminant Analizi Sınıflandırma Yüzdeleri

Yetenek			**Tahmin edilen grup üyelikleri**		**Toplam**
			Jimnastik	**Voleybol**	
Orjinal	**Sayı**	Jimnastik	240	34	274
		Voleybol	0	211	211
	%	Jimnastik	87,6	12,4	100,0
		Voleybol	0,0	100,0	100,0
Sınıflandırma başarısı %93,0					

Analiz sonucunda yapılan doğru sınıflandırma oranı %93 olmuştur.

5.5.4.2. Lojistik Regresyon Analizi: Tek Değişkenli Normalliğe Uygun, Çoklu Normalliği Sağlamayan; Varyansları Homojen, Kovaryansları Homojen Olmayan Veriler

DATA-3 verileri için yapılan diskriminant analizinin sonuçları ile karşılaştırmak için hiçbir varsayıma ihtiyaç duymayan lojistik regresyon analizi sonuçları elde edilmiştir.

Tablo 5.52 ile kurulacak olan lojistik regresyon modeline ilişkin bilgiler verilmiştir.

Tablo 5.52. DATA-3 Model Özeti

Adım	**-2 Log likelihood**	**Cox & Snell R^2**	**Nagelkerke R^2**
1	12,864	0,739	0,991

Tabloya göre bağımlı değişken ile bağımsız değişkenler arasında Cox and Snell'e göre yaklaşık %74, Nagelkerke'ye göre %99 ilişki olduğu söylenebilir.

Lojistik regresyon analizi ile kurulacak model için aşağıdaki Tablo 5.53'den yararlanılır.

Tablo 5.53. DATA-3 Lojistik Regresyon Analizi Denklem Değişkenleri

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Açık Atlama	8,207	10169,688	0,000	1	0,999	3668,118
	Smaç	65,436	10237,312	0,000	1	0,995	2621226844081953
	Krospas	-28,379	511,181	0,003	1	0,956	0,000
	Kondüsyon	0,204	0,274	0,554	1	0,457	1,226
	Sıçrama	-14,319	243,570	0,003	1	0,953	0,000
	Sabit	-373,405	244153,091	0,000	1	0,999	0,000
a. Variable(s) entered on step 1: AçıkAtlama, Smaç, Krospas, Kondüsyon, Sıçrama.							

Lojistik regresyon analizi sonuçlarına göre anlamlı değişken bulunmayıp, oluşturulan modele tüm değişkenler alınacaktır. Değişkenler arasında seçicilik olmayıp lojistik regresyon analizi adımsal yöntemler de birbirinden farklı değişkenleri modele almışlardır. Yöntemlerin modele aldıkları değişkenler aşağıda verilmiş olup diskriminant analizi ile aynı değişkenleri modele alan yöntemler forward LR ve forward conditionaldır.

Backward LR	smaç ve sıçrama
Backward Wald	açık atlama
Backward Conditional	smaç, krospas ve sıçrama
Forward LR	açık atlama ve sıçrama
Forward Wald	açık atlama
Forward Conditional	açık atlama ve sıçrama

Tablo 5.54'de lojistik regresyon analizinin sınıflandırma başarısı verilmiştir.

Tablo 5.54. DATA-3 Lojistik Regresyon Analizi Sınıflandırma Yüzdeleri

Gözlenen değerler			**Tahmin değerleri**		
			Yetenek		**Doğru sınıflama yüzdesi**
			Jimnastik	**Voleybol**	
Adım 1	Yetenek	Jimnastik	274	0	100,0
		Voleybol	2	209	99,1
	Toplam sınıflama yüzdesi				99,6

Tabloya göre lojistik regresyon analizinin DATA-3 verilerini kullanarak yaptığı sınıflandırma oranı %99,6'dır.

5.5.4.3. Probit Analizi: Tek Değişkenli Normalliğe Uygun, Çoklu Normalliği Sağlamayan; Varyansları Homojen, Kovaryansları Homojen Olmayan Veriler

Probit analizinde değişkenlerin etki sıralamasının nasıl olduğunu görmek için Tablo 5.55'den yararlanılır.

Tablo 5.55. DATA-3 Probit Analizi Parametre Tahminleri

Parametre		Tahmin	Std. Hata	Z	Sig.	%95 Güven aralığı	
						Alt sınır	Üst sınır
PROBIT	Açık Atlama	-0,864	1,063	-0,813	0,416	-2,946	1,219
	Smaç	1,885	1,036	1,819	0,069	-0,146	3,916
	Krospas	-0,492	0,529	-0,930	0,353	-1,529	0,545
	Kondüsyon	-0,008	0,066	-0,115	0,908	-0,136	0,121
	Sıçrama	-0,331	0,219	-1,511	0,131	-0,760	0,098
	Kesişim	-3,152	23,653	-0,133	0,894	-26,804	20,501

Probit analizi de lojistik regresyon analizi gibi anlamlı değişken bulmamıştır.

Aralarında güçlü korelasyon bulunan değişkenlerle yapılan analizlerin sınıflandırma oranları başarısının oldukça iyi, etkili değişkenleri bulmada zayıf kaldıkları görülmüştür. Diskriminant analizinde %93, lojistik regresyon analizinde ise %99,6 sınıflandırma başarısı olmuştur.

5.5.5. Kategorik Bağımsız Değişkenli Verilerle Analiz

Bu bölümün önceki başlıklarında normal dağılım, varyans homojenliği ve bağımsız değişkenler arasındaki ilişki durumlarına göre farklı yöntemlerin analiz sonuçları karşılaştırılmıştır. Bu karşılaştırmalarda nicel verilere sahip değişkenler kullanılmıştır. Oysa herhangi bir varsayım gerektirmeyen lojistik regresyon analizi ve probit analizinde bağımsız değişkenler nitel veriler de olabilir. Nitel veriye sahip değişkenler olması durumunda, bazı varsayımlar gerektiren bunun için de öncelikle nicel veri ile çalışılması gereken diskriminant analizinin sonuçları diğer iki yöntemle karşılaştırılmıştır. Dördüncü bölümde DATA-5 olarak adlandırılan; sigara içmeyi etkileyen faktörlerin araştırıldığı, hem nitel hem de nicel değişkenler içeren veriler ile yöntemlerin karşılaştırması yapılacaktır.

5.5.5.1. Diskriminant Analizi: Kategorik Bağımsız Değişkenli Veriler

Diskriminant analizi kurulacak olan fonksiyona alınacak değişkenleri aşağıdaki Tablo 5.56 ile belirlemiştir.

Tablo 5.56. DATA-5 Standartlaştırılmış Kanonik Diskriminant Fonksiyon Katsayıları

	Birinci fonksiyon
Cinsiyet	-0,580
Kanser Yaptığına İnanma	0,552
İnanca Aykırı Bulma	0,560
Kardeşin İçmesi	-0,339
Medeni Hal	0,309
Gelir	0,111
Yaş	0,026

Diskriminant analizi tüm değişkenleri analize dahil etmiştir.

Kurulacak olan model fonksiyonu için ise Tablo 5.57'deki katsayılar alınacaktır.

Tablo 5.57. DATA-5 Standartlaştırılmamış Kanonik Diskriminant Fonk. Katsayıları

	Birinci fonksiyon
Cinsiyet	-1,272
Kanser Yaptığına İnanma	2,675
İnanca Aykırı Bulma	1,146
Kardeşin İçmesi	-0,685
Medeni Hal	0,764
Gelir	0,000
Yaş	0,003
Cinsiyet	-2,520

Diskriminant analizi katsayılarına bakıldığında en çok etkiye sahip değişkenlerin sırasıyla; kanser yaptığına inanma, cinsiyet, inançlarına aykırı bulma, medeni hal ve kardeşin sigara içmesidir. Gelir ve yaş değişkenleri katsayıları sıfır ve sıfır sayılabilecek kadar küçük olup modele katkıları olmayacaktır.

Tablo 5.58 ile diskriminant analizinin DATA-5 verilerini sınıflandırmadaki başarısı verilmiştir.

Tablo 5.58. DATA-5 Diskriminant Analizi Sınıflandırma Yüzdeleri

Sigara İçme			**Tahmin Edilen Grup Üyelikleri**		**Toplam**
			Hayır	**Evet**	
Orjinal	**Sayı**	Hayır	123	15	138
		Evet	46	33	79
	%	Hayır	89,1	10,9	100,0
		Evet	58,2	41,8	100,0
Sınıflandırma başarısı % 71,9					

Diskriminant analizinin DATA-5 verilerini doğru sınıflandırma oranı %71,9 olmuştur.

5.5.5.2. Lojistik Regresyon Analizi: Kategorik Bağımsız Değişkenli Veriler

Lojistik regresyon analizinde DATA-5 ile kurulacak model için Tablo 5.59'daki katsayılardan yararlanılır.

Tablo 5.59. DATA-5 Lojistik Regresyon Analizinde Modele Alınan Değişkenler

		B	**S.E.**	**Wald**	**sd**	**Sig.**	**Exp(B)**
Adım 1[a]	Cinsiyet (1)	-1,154	0,396	8,502	1	0,004	0,315
	Kanser Yaptığına İnanma (1)	2,338	0,859	7,404	1	0,007	10,358
	İnanca Aykırı Bulma (1)	0,953	0,324	8,669	1	0,003	2,593
	Kardeşin İçmesi (1)	-0,576	0,307	3,510	1	0,061	0,562
	Medeni Hal (1)	0,699	0,450	2,409	1	0,121	2,011
	Gelir	0,000	0,000	0,244	1	0,621	1,000
	Yaş	-0,002	0,019	0,014	1	0,907	0,998
	Sabit	-0,783	0,791	0,981	1	0,322	0,457
a. Variable(s) entered on step 1: cinsiyet, kanser_yap_inanma, İnanc_aykırı_bulma, kardesin_icmesi, medeni_hal, gelir, yas.							

Lojistik regresyon analizi sonuçlarına göre cinsiyet, kanser yaptığına inanma, inançlarına aykırı bulma değişkenleri anlamlı bulunmuştur.

Tablo 5.60 ile lojistik regresyon analizinin DATA-5 verilerini sınıflandırmadaki başarısı verilmiştir.

Tablo 5.60. DATA-5 Lojistik Regresyon Analizi Sınıflandırma Yüzdeleri

Gözlenen değerler			Tahmin değerleri		
			Sigara İçme		Doğru Sınıflama Yüzdesi
			Hayır	Evet	
Adım 1	Sigara İçme	Hayır	123	15	89,1
		Evet	47	32	40,5
	Toplam sınıflama yüzdesi				71,4

Lojistik regresyon analizinin bu verileri doğru sınıflandırma oranı %71,4'dür.

5.5.5.3. Probit Analizi: Kategorik Bağımsız Değişkenli Veriler

Diskriminant ve lojistik regresyon analizi ile DATA-5 verileri için model kurarken değişkenlerin etki sıralamalarının aynı olduğu görülmüştür. Probit analizinde değişkenlerin etki sıralamasının nasıl olduğunu görmek için Tablo 5.61'den yararlanılır.

Tablo 5.61. DATA-5 Probit Analizi Parametre Tahminleri

Parametre		Tahmin	Std. Hata	Z	Sig.	%95 Güven aralığı	
						Alt sınır	Üst sınır
PROBIT	Cinsiyet	0,656	0,225	2,917	0,004	0,215	1,097
	Kanser Yaptığına İnanma	-1,357	0,479	-2,831	0,005	-2,296	-0,417
	İnanca Aykırı Bulma	-0,561	0,193	-2,907	0,004	-0,940	-0,183
	Kardeşin İçmesi	0,368	0,184	1,995	0,046	0,006	0,729
	Medeni Hal	-0,424	0,263	-1,610	0,107	-0,941	0,092
	Gelir	0,000	0,000	-0,563	0,573	0,000	0,000
	Yaş	-0,001	0,011	-0,091	0,928	-0,024	0,021
	Kesişim	0,849	0,670	1,267	0,205	0,179	1,518

Probit analizi lojistik regresyonla aynı değişkenleri anlamlı bulup farklı olarak bir değişkeni daha anlamlı bulmuştur. Bu değişken lojistik regresyon analizinde de anlamlı değişken olmaya oldukça yakın görülmektedir. Lojistik regresyon analizi ile %71,4 olarak bulunan doğru sınıflandırma oranı, diskriminant analizinde %71,9 olarak bulunmuştur. Hem diskriminant analizi hem de lojistik regresyon analizinin farklı adımsal yöntemleri ile yapılan analizlerde model kurmak için anlamlı bulunan değişkenler kanser yaptığına inanma, cinsiyet, inançlarına aykırı bulma ve kardeşin sigara içmesidir.

6. SINIFLANDIRMA MODELLERİNDE KARŞILAŞTIRMA

6.1. Faktör Analizi

1930'lu yıllarda çalışılan faktör analizi özellikle sosyal alanlarda kullanılan, ilk uygulamaları psikoloji alanında yapılan ve bilgisayarların gelişimi ile farklı alanlarda da kullanılan çok değişkenli istatistik yöntemdir. Faktör analizi başlı başına bir yöntem olduğu gibi temel işlevi dışında diğer bazı yöntemlerin uygulanmasında da yararlanılan yöntemdir. Farklı amaç ve durumlar için kullanılan faktör analizine değişik tanımlamalar yapılmıştır. Faktör analizi genellikle çok sayıdaki değişkenin daha az sayıdaki temel değişkenle incelenip incelenemeyeceği merak edildiğinde kullanılmaktadır. Çok sayıdaki değişkenin birkaç başlık altında toplanıp toplanmayacağına karar veren, belli başlıklarda toplanan verileri yeniden adlandıran bir yöntemdir. Böylece faktör analizi ile boyut indirgeyip değişken sayısı azaltıldığı gibi, bu değişkenler sınıflandırılmış da olmaktadır (Alpar, 2017). Birbiri ile ilişkili değişkenleri aynı gruba toplayarak az sayıda yeni ve birbirleriyle ilişkisiz değişkenler bulmayı amaçlar (Toktay, 2017). Faktör analizi, birbiriyle bağlantılı çok sayıda değişkeni bir gruba toplayarak, daha az sayıda anlamlı yeni değişkenler bulmayı amaçlayan istatistik yöntem olarak da tanımlanabilir (Tavşancıl, 2014). Faktör analizinde amaç değişkenler arasındaki korelasyonları göz önüne alarak mümkün olan en az sayıda değişkenler grubu belirlemektir. Böylece değişkenlerin temelinde yatan ortak ilişkiler analiz edilebilir. Faktör analizi ile elde edilen faktör olarak adlandırılan yeni değişkenler diskriminant analizi, regresyon ve korelasyon analizi gibi yöntemler için değişken olarak kullanılabilir (Albayrak, 2006).

Bu tanım ve amaçlardan faktör analizinde değişkenler arasında bağımlı ve bağımsız değişken ayrımı olmadığı anlaşılmaktadır. Aralarında az veya çok ilişki bulunan tüm değişkenler eş anlı olarak analiz edilmektedir. Diğer bazı yöntemlerde olduğu gibi bağımlılık yapısı araştırılmamaktadır. Faktör analizi yapılırken öncelikle verilerin faktörlenebilecek bir yapıda olup olmadığına karar verilir. Çok değişkenli varsayımları sağlayıp sağlamadığına bakılır.

Faktör çıkarma yöntemlerinden biri ile faktör yükleri matrisi çıkarılır. Bunun için en sık kullanılan yöntemler temel bileşenler analizi ve en çok olabilirlik yöntemleridir. Bunun devamında ise faktör döndürme yöntemlerinden biri ile faktörler daha iyi

yorumlanacak hale gerilmeye çalışılır. Daha sonra özdeğerlerin incelenmesi ile analiz sonucunda kaç faktörün dikkate alınacağı belirlenmeye çalışılır. Son olarak da faktörlerin yani temel değişkenler altında toplanan yeni değişkenlerin isimlendirilmesi, yorumlanması ve araştırmanın sonuç bölümüne geçilir. Faktör yükleri temel bileşenler yöntemi, temel eksen faktörleştirme yöntemi ve görüntü faktörleştirme yöntemi gibi yöntemlerle elde edilebilir. Bu durumda veri matrisinin kendisi yerine veri matrisinden elde edilen varyans-kovaryans matrisi veya korelasyon matrisi gibi ikincil verilerden faydalanılır. Faktör analizinde, çok sayıda olmamak koşuluyla, kukla değişkenler kullanılabilmektedir (Albayrak, 2005).

Büyüköztürk (2002), faktör analizi temel kavramlarını ve ölçek geliştirmede kullanımı üzerine çalışmıştır. Işık ve ark. (2004), Türkiye ekonomisindeki finansal krizlere faktör analizi uygulamışlardır. Öven ve Pekdemir (2005), ofislerin kira bedellerinde etkili olan değişkenleri faktör analizi ile incelemişlerdir. Doğan ve Başokçu (2010), faktör analizi ve kümeleme analizini karşılaştıran bir çalışma yapmışlardır. Yaşlıoğlu (2017), sosyal bilimlerde faktör analizi ve geçerlilik çalışmıştır.

6.2. Kümeleme Analizi

Küme kelimesi denilince birbirine benzer nesnelerin oluşturduğu topluluk akla gelmektedir. Kümeleme analizi, benzer gözlemleri bazen de benzer değişkenleri homojen alt gruplara ayırmak için kullanılan çok değişkenli istatistik analiz tekniğidir. Benzer gözlemleri aynı kümelere toplayarak boyut indirgeyen kümeleme analizine sağlık, ziraat, ekonomi, anket çalışmaları vs. birçok alanda sıklıkla başvurulmaktadır. Kümeleme analizi, genellikle gözlemleri kümelere ayırmak için kullanılsa da değişkenleri kümelemek için de kullanılabileceği belirtilmektedir. Ancak değişkenlerin kümelere ayrılmasında çoğunlukla faktör analizi kullanılır. Kümeleme analizi ile aşırı değerler de belirlenebilmektedir. Analiz sonucunda hiçbir kümeye dahil olamayan tek başına bir küme oluşturmuş gibi sonuçlanan bir gözlemin aşırı değer olduğu düşünülebilir.

Bu analizin başarısı için aynı kümede toplanan nesnelerin kendi aralarında mümkün olduğunca benzer, kümeler arasındaki farkın ise mümkün olduğunca farklı olması istenir. Gözlemlerin gruplandırılmasında kullanılan kümeleme analizi diskriminant analizine de benzetilmektedir. Ancak kümeleme analizinde, analiz

sonucunda elde edilecek gruplar, diskriminant analizinde çalışmanın başlangıcında yapılır. Bu daha önce kesin bilinen öldü-yaşıyor gibi gruplar veya araştırmacının kendi tecrübe ve birikimi ile belirlediği gruplardır. Bu yönüyle birbirine alternatif yöntemler olamazlar.

Kümeleme analizi sonucunda elde edilecek küme sayısını belirlemede standart bir yaklaşım bulunmamaktadır. Bu nedenle birden fazla sonuç da elde edilebilir. Herhangi bir gözlemin ait olduğu küme, kullanılan yönteme ve küme sayısına göre değişebilir. Bu nedenle öncelikle beklenen, analiz sonucunda kümeler arasında anlamlı farklar bulunmalı ve oluşan kümeler kuramsal açıdan geçerli olmalıdır. Sonuçların en doğru değerlendirilebilmesi için çalışma konusu olan, alan bilgisine iyi hakim olmak gereklidir. Kümeleme analizi diğer bazı istatistik yöntemlerde olduğu gibi örneklemden elde edilen sonuçlara göre evren hakkında tahminde bulunmaya imkan sağlayacak bir yöntem değildir.

Kümeleme analizi için önemli olan ön değerlendirme değişkenler arasında çoklu bağlantı olup olmadığıdır. Diğer bazı yöntemlerde aranan normallik, varyans homojenliği gibi varsayımlar kümeleme analizi için çok önemli değildir (Alpar, 2017). Kümeleme analizinde diğer bazı yöntemlerde olduğu gibi bağımlı, bağımsız değişken ayrımı bulunmamaktadır. Kümeleme analizi gözlemlerin tüm değişkenler üzerindeki ölçülen değerlerini hesaplayarak gözlemleri kümelere ayırmaktadır. Gözlemler arasındaki benzerlikleri belirlemek için korelasyon ölçüleri, uzaklık ölçüleri veya nitelik veriler için benzerlik ölçülerini kullanmaktadır. Sınıflama başarılı ise aynı küme içindeki gözlemler geometrik olarak birbirlerine yakın, farklı kümeler arasındakiler ise uzaktırlar (Kalaycı, 2010)

Çakmak (1999), kümeleme analizinin geçerliliği ve sonuçlarının değerlendirilmesini incelemiştir. Çakmak ve ark. (2005), kümeleme analizi ile illerin kültürel yapılarına göre sınıflandırılması ve değişimlerini inceleme çalışması yapmışlardır. Ersöz (2009), sağlık göstergeleri üzerine çalışırken kümeleme ve ayırma analizlerinden faydalanmıştır. Topak (2010), imalat sanayinde faaliyet gösteren firma risklerinin belirlenmesi çalışmasında kümeleme analizi üzerinde çalışmıştır. Altun Ada (2011), AB ülkeleri ve Türkiye'nin sürdürülebilir kalkınma açısından değerlendirmesinde kümeleme analizini kullanmıştır. Çelik (2013), kümeleme analizi ile Türkiye'de illerin sağlık göstergelerine göre sınıflandırılma çalışması yapmıştır.

6.3. Faktör ve Kümeleme Analizlerinin Karşılaştırma Uygulaması

Üniversiteye giriş sınavlarına hazırlanan 100 öğrencinin yapmış oldukları netlere göre benzer derslerin gruplandırılmak istendiği bir çalışma ele alınsın. Benzer derslerin aynı grupta toplanma durumları faktör ve kümeleme analizleri yapılarak iki yöntem arasında karşılaştırma yapılacaktır.

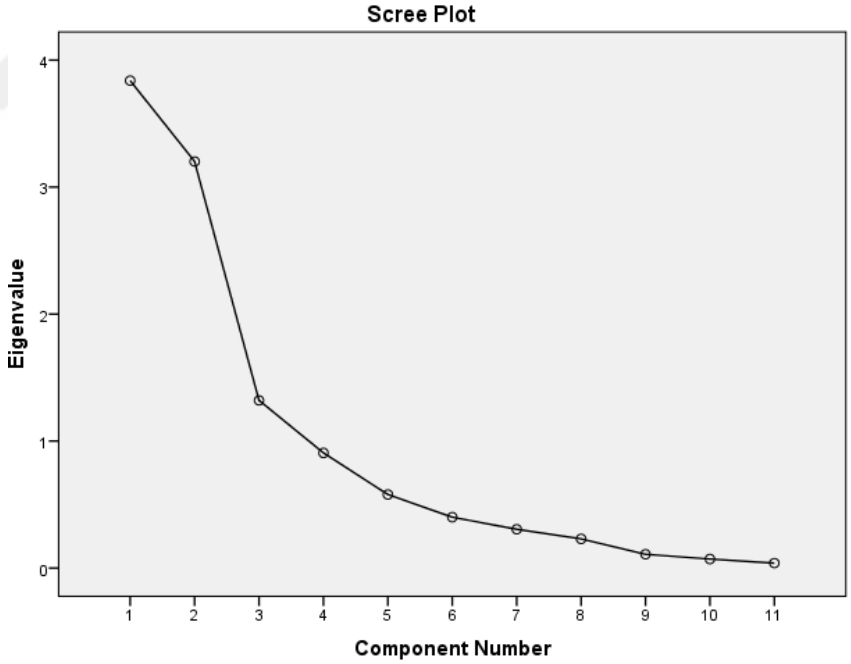
Tablo 6.1. Faktör Matrisi

	Faktörler		
	1	2	3
Türkçe	0,936		
Dil	0,825		
Cebir	-0,758		
Geometri	-0,718		
Coğrafya		0,955	
Felsefe		0,950	
Tarih		0,939	
Din		0,762	
Fizik			0,952
Kimya			0,911
Biyoloji			0,733

Tablo 6.1'deki analiz sonuçlarına göre Türkçe, dil, cebir ve geometri dersleri aynı faktör altında toplanmıştır. İkinci faktörde tarih, coğrafya, felsefe ve din dersi toplanmıştır. Fizik, kimya, biyoloji ise üçüncü faktörü oluşturmuştur.

Tablo 6.2. Üç Faktör ile Açıklanan Varyans Yüzdeleri

Faktör	İlk özdeğer			Elde edilen yüklerin kareleri toplamı			Döndürülmüş yüklerin kareleri toplamı
	Toplam	Varyans yüzdeleri	Birikimli yüzde	Toplam	Varyans yüzdeleri	Birikimli yüzde	Toplam
1	3,839	34,896	34,896	3,839	34,896	34,896	3,215
2	3,201	29,103	63,998	3,201	29,103	63,998	3,368
3	1,320	12,002	76,000	1,320	12,002	76,000	2,872
4	0,907	8,242	84,242				
5	0,579	5,264	89,507				
6	0,401	3,641	93,148				
7	0,306	2,778	95,926				
8	0,230	2,088	98,014				
9	0,108	0,986	98,999				
10	0,071	0,645	99,644				
11	0,039	0,356	100,000				



Şekil 6.1. Yamaç Grafiği

Yukarıdaki Tablo 6.2 ve Şekil 6.1 incelendiğinde dördüncü faktörün total değerinin 1 değerine yakın olduğu ve üç faktörle toplam varyansın %76'sı açıklanırken dört faktörle bu oranın %84 olduğu görülmektedir. Yine yamaç grafiği incelendiğinde

de faktör sayısının dört olarak da belirlenebileceği görülür. Faktör sayısı dört olarak belirlenip, analiz tekrar yapıldığında aşağıdaki sonuçlar elde edilir.

Tablo 6.3. Faktör Matrisi

	Faktörler			
	1	2	3	4
Dil	0,905			
Türkçe	0,839			
Felsefe		0,976		
Coğrafya		0,968		
Tarih		0,968		
Din		0,652		0,462
Fizik			0,966	
Kimya			0,943	
Biyoloji			0,659	0,325
Cebir				0,766
Geometri				0,735

Tablo 6.3'e göre faktör sayısı dört olarak seçilip analiz yapılması durumunda, derslerin gruplandığı dört faktörün aşağıdaki gibi olduğu görülebilir.

1. faktör: dil ve Türkçe,
2. faktör: felsefe, coğrafya, tarih ve din
3. faktör: fizik, kimya, biyoloji
4. faktör: cebir ve geometri

Birinci faktör için dil grubu, ikinci faktör için sosyal bilimler grubu, üçüncü faktör için fen bilimleri grubu, dördüncü faktör için matematik grubu isimlendirilmesi yapılabilir.

Benzer olan sınavları gruplama işlemi faktör analizine benzer olarak kümeleme analizi ile yapılması durumunda elde edilen sonuçlar aşağıda verilmiştir. Faktör analizinde olduğu gibi öncelikle derslerin üç kümeye ayrılması durumunda oluşacak gruplar için Tablo 6.4'den yararlanılır.

Tablo 6.4. Üç Kümeli Gruplandırmada Küme Üyelikleri

Sıra	Dersler	Küme	Mesafe
1	Fizik	3	15,583
2	Kimya	3	15,020
3	Biyoloji	3	13,301
4	Cebir	1	0,000
5	Geometri	3	41,248
6	Tarih	3	13,745
7	Coğrafya	3	12,611
8	Felsefe	3	13,264
9	Din	3	12,929
10	Türkçe	2	0,000
11	Dil	3	41,194

Birinci kümeye sadece cebir dersi, ikinci kümeye sadece Türkçe dersi alınmışken, diğer tüm dersler aynı kümeye toplanmıştır. Tablo 6.5'de bu durumu her bir kümeye atanan ders sayıları ile vermektedir.

Tablo 6.5. Üç Kümeli Gruplandırmada Her Kümeye İsabet Eden Gözlem Sayısı

Küme	**1**	1,000
	2	1,000
	3	9,000
Atanan		11,000
Atanmayan		0,000

Tabloya göre de birinci ve ikinci kümeye sadece 1'er ders, üçüncü kümeye ise 9 ders atanmıştır.

Dört grup olmaları durumundaki faktör ve kümeleme analizlerini karşılaştırmak için; faktör analizinde olduğu gibi küme sayısı dört olarak seçilip analizler tekrar yapıldığında kümeler Tablo 6.6'daki gibi oluşmaktadır.

Tablo 6.6. Dört Kümeli Gruplandırmada Küme Üyelikleri

Sıra	Dersler	Küme	Mesafe
1	Fizik	1	12,275
2	Kimya	1	13,103
3	Biyoloji	1	14,411
4	Cebir	4	0,000
5	Geometri	1	31,308
6	Tarih	2	10,874
7	Coğrafya	2	10,660
8	Felsefe	2	10,883
9	Din kültürü	2	11,775
10	Türkçe	3	0,000
11	Dil bilgisi	2	35,069

Fizik, kimya, biyoloji ve geometri dersleri birinci kümede toplanmıştır. Tarih, coğrafya, felsefe, din ve dil dersleri ikinci kümeyi oluşturmaktadır. Üçüncü kümede sadece Türkçe dersi, dördüncü kümede ise sadece cebir dersi yer almaktadır. Tablo 6.7'de bu durumu her bir kümeye atanan ders sayıları ile vermektedir.

Tablo 6.7. Dört Kümeli Gruplandırmada Her Kümeye Atanan Gözlem Sayısı

Küme	**1**	4,000
	2	5,000
	3	1,000
	4	1,000
Atanan		11,000
Atanmayan		0,000

Tabloya göre birinci kümeye 4, ikinci kümeye 5, üçüncü ve dördüncü kümelere 1'er ders atanmıştır.

Mevcut eğitim sistemine göre derslerin gerek üç gerekse dört gruba ayrılmasında faktör analizinin daha gerçekçi bir gruplandırma yaptığı düşünülebilir. Üç faktöre ayrılan eğitim grupları isimlendirilecek olursa Türkçe, dil, cebir ve geometri derslerine eşit ağırlıklı; tarih, coğrafya, felsefe ve din derslerine sözel ağırlıklı ve fizik, kimya, biyoloji derslerine fen ağırlıklı sınıf öğrencileri denilebilir. Oysa kümeleme analiziyle dersler üç kümeye ayrıldığında, mevcut güncel eğitim sistemi de göz önüne alındığında iyi bir gruplandırma yapılamadığı söylenebilir. Çünkü kümeleme analizi, cebir ve Türkçe derslerini ayrı birer küme kabul etmiş, diğer tüm dersleri üçüncü kümeye

toplamıştır. Derslerin dört gruba ayrılması halinde de faktör analizinin, eğitim sistemindeki duruma daha uygun sonuçlar verdiği söylenebilir. Türkçe ve dil dersleri için dil sınıfı; felsefe, coğrafya, tarih ve din dersleri için sosyal bilimler sınıfı; fizik, kimya, biyoloji dersleri için fen bilimleri sınıfı ve cebir, geometri dersi için matematik sınıfı öğrencileri denilebilir. Dört küme ile yapılan kümeleme analizinin, üç küme ile yapılan kümeleme analize göre eğitim sistemine daha uyumlu olduğu, ancak faktör analizine göre yine zayıf kaldığı söylenebilir. Çünkü kümeleme analizi, Türkçe ve cebir derslerini yine ayrı birer küme olarak değerlendirmiş; fizik, kimya, biyoloji ve geometri derslerini bir kümeye; tarih, coğrafya, felsefe, din ve dil derslerini de ayrı bir kümeye toplamıştır. Dört kümeli kümeleme analizinin, üç kümeli analize göre bir iyileşme gösterdiği söylenebilir. Türkçe ve cebir derslerini yine ayrı birer küme olarak değerlendirse de aynı kümeye topladığı üçüncü kümeyi ikiye ayırırken fizik, kimya, biyoloji ve geometri dersleri için sayısal sınıfı; tarih, coğrafya, felsefe, din ve dil dersleri için sözel sınıfı öğrencileri denilebilir.

7. SONUÇ

Bu çalışmanın gerek içeriği gerekse sonuçları değerlendirilirken istatistiğin deterministik değil stokastik ilişkilerle ilgilendiği gerçeği göz önünde bulundurulmadır. Burada elde edilen sonuçlar ile genel geçer kurallar elde edilmemiş olup, bu çalışmada kullanılan veriler ve simülasyonlar ile elde edilen sonuçlar paylaşılmıştır. Elde edilen sonuçlara göre herhangi iki veya daha fazla yöntemin birbiriyle tamamen uyumlu olduğu veya birinin diğerlerinden daha üstün olduğu sonucu çıkarılamaz. Benzer çalışmaların yapılması ve bu çalışma sonuçları ile karşılaştırılması halinde daha çok değer kazanacağı düşünülmektedir. Herhangi bir konuda istatistik analiz sonuçlarını değerlendirmek için sadece konunun uzmanı olmak yeterli olmadığı gibi sadece iyi bir istatistik bilgisine sahip olmanın da yeterli olmadığı unutulmamalıdır. Bu ikisinin bir arada olması her zaman mümkün olmasa da konunun uzmanları ile yapılacak ortak çalışmalardan daha verimli sonuçlar alınacağı düşünülmektedir. Kullanılacak olan yönteme karar verirken araştırma konusunun ve birimlerinin de dikkate alınması önemlidir. Yanlış bir yöntem veya karar sonucunda olumsuz etkilenecek olan araştırma birimleri de yöntem seçmede hassasiyet nedenlerinden biri olabilir. Örneğin sağlık alanında yapılan bir çalışmada kobay kullanılmasında her ne kadar etik kurallara uyulması gerekse de olumsuz etkileri doğrudan insanlara yansıyacak çalışmalarda çok daha titiz davranılması gerekmektedir.

Bu çalışmada tek değişkenli normallik değerlendirmeleri yapılırken literatürde en çok kullanılan yöntemler kullanılmış olup bu yöntemlerin, değişkenleri normalliğe kabul etme derecelerinde farklılıklar bulunduğu gözlemlenmiştir. Aynı değişkenlerin çok değişkenli normalliklerine de bakılarak tek ve çok değişkenli normallikler arasındaki ilişkiye de bakılmıştır.

DATA-1 verileri tek değişkenli normal dağılım için önerilen ve bu çalışmada yararlanılan görsel ve analitik yöntemlerin tümüne göre ayrı ayrı tek değişkenli normalliği sağlamıştır. Bu veriler çok değişkenli normallik şartını da sağlamışlardır. Sadece bu veriler ile yapılacak olan değerlendirmeye göre tek değişkenli normalliği sağlayan verilerin her zaman çok değişkenli normallik şartını da sağlayacakları veya bunun tersi her zaman söylenemez. Bu çalışmanın DATA-2 olarak kullanılan verilerinde bu durumun olmadığı görülmüştür.

DATA-2 verileri tek değişkenli normallik değerlendirmelerine göre tek değişkenli normallik varsayımını sağlamadıkları halde çok değişkenli normallik şartını sağlamışlardır. O halde değişkenlerin tek değişkenli normalliği sağlamamaları çok değişkenli normalliği de sağlamayacakları anlamına gelmeyebilir. Veya çok değişkenli normal dağılıma uygun bulunan verilere ait değişkenlerin ayrı ayrı tek değişkenli normalliği de sağlayacakları söylenemeyebilir.

DATA-3 verileri tek değişkenli normallik değerlendirmelerine göre tek değişkenli normallik varsayımını sağladığı kabul edilebilecek değişkenler iken çok değişkenli normallik şartını sağlamadığı görülmüştür. Bu durum değişkenlerin tek değişkenli normalliği sağlamaları durumunda çok değişkenli normalliği de sağlayacakları fikrine uygun bulunmamıştır.

Yöntemlerin, değişkenleri normalliğe kabul etme derecelerindeki farklılıklar DATA-1, DATA-2 ve DATA-3'de gözlemlenmiş bu durum DATA-4 verileri ile tekrar incelenerek yöntemlerin aynı değişkeni normal dağılıma uygun kabul etme durumlarının nasıl olduğu anlaşılmaya çalışılmıştır. Bu yöntemlerden, çarpıklık ve basıklık değerlerinin belli aralıkta olmasını öneren yönteme göre normallik şartı genellikle sağlanmaktadır. Bu yönteme benzer olan, çarpıklık ve basıklık değerlerinin kendi standart hatalarına bölünmesinin belli bir aralıkta olmasını öneren yöntem ise sadece çarpıklık ve basıklık değerlerine bakan yöntemden farklı olarak bazı değişkenlerin normal dağılıma uygun kabul etme aralığının uç değerlerine yaklaştırarak normal dağılıma uygun bulmaktadır. Yani normallikten sapmaya yakın bulmaktadır. Her iki yöntemde de normal kabul etme için standart bir aralık bulunmaması bir dezavantaj olduğu gibi bu yöntemler çarpıklık ve basıklığı ayrı ayrı değerlendirmekte olup, bunlardan birinin sağlanıp, birinin sağlanmaması durumunda net bir sonuç vermemektedir. Aynı değişken için basıklık ve çarpıklıktan birinin belirlenen aralıkta kalması diğerinin aralığın dışında kalması durumlarında ayrıca bir değerlendirme gerektirmektedir.

Tek Örnek Kolmogorov-Simirnov testi, çarpıklık ve basıklığı ayrı sonuçlar olarak vermeyip paket programların anlamlılık düzeyi olarak verdiği tek sonuç ile karar vermeyi kolaylaştırmaktadır. Bu yöntem yukarıdaki iki yöntem ile aynı değişkenler için karşılaştırılmıştır. Yukardaki yöntemler ile değişkenlerin normal dağılıma uygun kabul etme aralığının uç değerlerine yaklaşarak normal dağılıma uygun bulunan değişkenler

yani normallikten sapmaya yakın bulunan değişkenler Tek Örnek Kolmogorov-Simirnov testi ile normal dağılıma uygun olmadığı görülmüştür.

Benzer şekilde düzeltilmiş Kolmogorov-Smirnov (Lilliefors) testi de Tek Örnek Kolmogorov-Simirnov testinin normal kabul ettiği bazı değişkenlerin normal dağılıma uygunluğunu reddedebilmektedir. Shapiro-Wilk istatistiği de düzeltilmiş Kolmogorov-Smirnov (Lilliefors) testinin normal kabul ettiği bazı değişkenlerin normalliğini reddedebilmektedir.

Normallik değerlendirmesinde bir değişkenin en çok normalliğe uygunluk lehine sonuç veren yöntemin çarpıklık ve basıklık değerlerini dikkate alan yöntemler olduğu, anlamlılık düzeylerine bakarak değerlendirme yapan yöntemler arasında ise Tek Örnek Kolmogorov-Simirnov testi olduğu, en az normalliğe uygunluk lehine sonuç veren yöntemin ise Shapiro-Wilk testi olduğu söylenebilir. Bu durumda Shapiro-Wilk testinin normale yakın olan değişkenlerin de normalliğini reddedebildiği, Tek Örnek Kolmogorov-Smirnov testinin ise normalden uzak verileri de normal kabul edebileceği söylenebilir.

Normal dağılıma uyan veya uymayan nicel veriler hatta nitel verilerin korelasyon katsayısı karşılaştırmalarında bu çalışmada kullanılan veriler için Pearson, Spearman ve Kendall's tau yöntemleri benzer sonuçlar vermiştir.

Diskriminant analizi, lojistik regresyon analizi ve probit analizi karşılaştımaları şu özellikteki veriler için yapılmıştır:

- Tek ve çok değişkenli normal dağılıma uygun, varyans ve kovaryans homojenliği olmayan veriler
- Tek değişkenli normal olmayıp, çoklu normalliği sağlayan; varyans ve kovaryans homojenliği olan veriler
- Tek değişkenli normal olmayıp, çoklu normalliği sağlayan; varyansları homojen, kovaryansları homojen olmayan veriler
- Tek değişkenli normalliğe uygun, çoklu normalliği sağlamayan; varyansları homojen, kovaryansları homojen olmayan veriler
- Kategorik bağımsız değişkenli veriler.

Yukarıda verilen tek ve çok değişkenli normalliğe uygunluk ile varyans, kovaryans homojenliği şartlarına uyma durumlarına göre farklılık gösteren veriler için diskriminant, lojistik regresyon ve probit analizi sonuçları karşılaştırılmıştır.

Yöntemlerin modele aldıkları değişkenler arasında ve doğru sınıflandırma yüzdeleri arasında önemli bir fark bulunamamıştır.

Daha çok değişkenleri benzer gruplara ayırmak için kullanılan faktör analizi ile gözlemleri benzer gruplara ayırmak için kullanılan kümeleme analizinin sonuçlarının benzer olmadığı söylenebilir.

Bu çalışma sonucunda istatistiksel yöntemlerin sonuçları değerlendirilirken, tek bir yönteme göre karar vermek yerine farklı yöntemlerin sonuçlarını birlikte değerlendirmenin ve çalışmanın önemi dikkate alınarak değerlendirme yapmanın daha uygun olacağı düşünülmektedir.

KAYNAKÇA

Akdamar, E., (2018)., *Akıllı Kentlere İlişkin ISO 37120 Standardı Göstergelerinin Çok Değişkenli İstatistiksel Tekniklerle İrdelenmesi*, (Doktora Tezi), Uludağ Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, İstatistik Bilim Dalı, Bursa.

Akdeniz F., "New Trends And Developments In Statistics", *Social Sciences Research Journal*, 4/2015, ss. 1–11.

Akın, F., (2002), Ekonometri, (1. baskı), Ekin Yayınevi, Bursa.

Akkaya, Ş. ve Pazarlıoğlu M. V., (1998), Ekonometri II, (2. baskı), Erkam Matbaası, İstanbul.

Akkaya M., Kantar L., "Finansal Krizlerin Tahmininde Öncü Göstergelerin Logit-Probit Model İle Analizi: Türkiye Uygulaması", *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 14/2018, (3), ss. 575–590.

Aktaş, C. ve Erkuş, O., "Lojistik Regresyon Analizi ile Eskişehir'in Sis Kestiriminin İncelenmesi", *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Güz 2009, (16), ss. 47–59.

Albayrak, S., (2003), *Türkiye'de İllerin Sosyo-Ekonomik Gelişmişlik Düzeylerinin Çok Değişkenli İstatistik Yöntemlerle İncelenmesi*, (Doktora Tezi), İstanbul Üniversitesi, İstanbul.

Albayrak, A. S., (2006), Uygulamalı Çok Değişkenli İstatistik Teknikleri, (1. baskı), Asil Yayınları, Ankara.

Albayrak, A. S., "Türkiye'de İllerin Sosyoekonomik Gelişmişlik Düzeylerinin Çok Değişkenli İstatistik Yöntemlerle İncelenmesi", *Zonguldak Karaelmas Üniversitesi Sosyal Bilimler Dergisi*, 2005/1 (1), ss. 153–177.

Alpar, R., (2017), Uygulamalı çok değişkenli istatistiksel yöntemler, (4. baskı), Detay Yayıncılık, Ankara.

Altıntaş T., Duru M. N., "Lojistik Regresyon ve Probit Model ile Marka Bağımlılığı Araştırması", *Anadolu BİL Meslek Yüksekokulu Dergisi*, 2007/2, (7), ss. 56–65.

Altun Ada, A., "Kümeleme Analizi İle Ab Ülkeleri Ve Türkiye'nin Sürdürülebilir Kalkınma Açısından Değerlendirilmesi", *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, Nisan 2011, (29), ss. 319–332.

Anderson, J. A., "Separate Sample Logistic Diskrimination", *Biometrica*, April 1972/50, (1), ss. 19–35.

Arslan, O., "Su Kalitesi Verilerinin CBS ile Çok Değişkenli İstatistik Analizi (Posuk Çayı Örneği)", *Jeodezi, Jeoinformasyon ve Arazi Yönetimi Dergisi*, 2008/2, (99), ss. 5–11.

Atakan, C., Karabulut, İ., "Derinliğe Dayalı Diskriminasyon", *Selçuk Üniversitesi Fen Edebiyat Fakültesi Fen Dergisi*, 2003, (22), ss. 53–63.

Atkinson, A. C., "Atest of the Linear Logistic and Bradley-Terry Models", *Biometrica,* 1972/59, (1), ss. 37–42.

Başarır, G., (1990), *Çok Değişkenli Verilerde Ayrımsama Sorunu Ve Logistik Regresyon Analizi*, (Uygulamalı İstatistik Doktora Tezi), Hacettepe Üniversitesi, 1-36, Ankara.

Bayata H. F., Hattatoğlu, F., "Yapay Sinir Ağları Ve Çok Değişkenli İstatistik Yöntemlerle Trafik Kaza Modellemesi", *EÜFBED - Fen Bilimleri Enstitüsü Dergisi*, 2010/3, (2), ss. 207–219.

Begg, C. B., Gray, R., "Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions", *Biometrica,* 1982/71, (1), ss. 11–18.

Bircan, H., (2004), "Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama", *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi,* 2004, (8), ss. 185–208.

Biçkici, B., (2007), *Çok Değişkenli Varyans Analizi Ve Çoklu Doğrusal Regresyon Analizinin Uygulamalı Olarak Karşılaştırılması*, (Yüksek Lisans Tezi), Atatürk Üniversitesi, Erzurum.

Bonney, G. E., "Logistic Regression for Dependent Binary Observations", *Biometrica,* 1987/43, ss. 951–973.

Burmaoğlu, S., Oktay, E. ve Özen, Ü., "Birleşmiş Milletler Kalkınma Programı Beşeri Kalkınma Endeksi Verilerini Kullanarak Diskriminant Analizi ve Lojistik Regresyon Analizinin Sınıflandırma Performanslarının Karşılaştırılması", *Savunma Bilimleri Dergisi*, 2009/8, (2), ss. 23–49.

Büyüköztürk, Ş., "Faktör Analizi: Temel Kavramlar ve Ölçek Geliştirmede Kullanımı", *Kuram ve Uygulamada Eğitim Yönetimi*,. Güz 2002, (32), ss. 470-483.

Cain, M. K., Zhang, Z., and Yuan, K. H., "Univariate and Multivariate Skewness and Kurtosis for Measuring Nonnormality: Prevalence, Influence and Estimation", *Behavior Research Methods*, 2017/*49,* (5), ss. 1716–1735.

Can, S., (2011), *Bazı Çok Değişkenli İstatistiksel Teknikler Arasındaki İlişkinin İncelenmesi Ve Uygulamaları*, (Yüksek Lisans Tezi), Çukurova Üniversitesi, Adana.

Carroll, R. J., Spiegelman, C. H., Lan, G. K. K., Bailey, K. T. and Abbott, R. D., "On Error-Invariables for Binary Regression Models", *Biometrica,* 1982/71, (1), ss. 19–25.

Cebeci, İ., "Krizleri İncelemede Kullanılan Nitel Tercih Modelleri: Türkiye İçin Bir Probit Model Uygulaması: (1988-2009)", *İstanbul Üniversitesi İktisat Fakültesi Mecmuası,* 2012/62, (1), ss. 127–146.

Coşkun, S., Kartal, M., Coşkun, A. ve Bircan, H., "Lojistik Regresyon Analizinin İncelenmesi Ve Dişhekimliğinde Bir Uygulaması", *Cumhuriyet Üniversitesi Diş Hekimliği Fakültesi Dergisi,* 2004/7, (1), ss. 41–50.

Cox, D. R., Snell, E. J., (1989), Analysis of binary DATA, Chapman and Hall, 236, London, UK.

Çakmak, Z., "Kümeleme Analizinde Geçerlilik Problemi ve Kümeleme Sonuçlarının Değerlendirilmesi", *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi,* Kasım 1999, (3), ss. 187–205.

Çakmak, Z., Uzgören, N., Keçek, G., "Kümeleme Analizi Teknikleri ile İllerin Kültürel Yapılarına Göre Sınıflandırılması ve Değişimlerinin İncelenmesi", *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi,* 2005, (12).

Çelik, Ş., "Kümeleme Analizi ile Sağlık Göstergelerine Göre Türkiye'deki İllerin Sınıflandırılması", *Doğuş Üniversitesi Dergisi,* 2013/14, (2), ss. 175–194.

Çelik, Ş., İnci, H., Şengül, T. ve Söğüt, B., "Diskriminant Analizi ile Bıldırcın Yumurtalarında Bazı Kalite Özellikleri ile Tüy Rengi Arasındaki İlişkinin İncelenmesi", *Tekirdağ Ziraat Fakültesi Dergisi,* 2015/12, (3), ss. 47–56.

Çiftçi, M., "Kalkınma Göstergesi Olarak Ortalama Yaşam Beklentisine Göre Türkiye' Nin AB İçindeki Konumu: Kriterler Ve Çok Değişkenli İstatistik Uygulamaları", *Ekonometri ve İstatistik,* 2008, (7), ss. 51–87.

Çokluk, O., (2010), "Logistic Regression: Concept and Application", *Educational Sciences: Theory and Practice,* Summer 2010/10, (3), ss. 1357–1407.

Demir E, Saatçioğlu Ö, İmrol F., "Uluslararası Dergilerde Yayımlanan Eğitim Araştırmalarının Normallik Varsayımları Açısından İncelenmesi", *Current Research in Education,* 2016/2, (3), ss. 130–148.

Dobson, A. J., (1990), An introduction to Generalized Linear Models, (First Edition), Chapman and Hall, 173, New York-USA.

Doğan, N., Başokçu, O., "İstatistik Tutum Ölçeği İçin Uygulanan Faktör Analizi ve Aşamalı Kümeleme Analizi Sonuçlarının Karşılaştırılması", *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi,* Kış 2010/1, (2), ss. 65–71.

Erçetin, Y., (1993), "Diskriminant Analizi ve Bankalar Üzerine Bir Uygulama", *Türkiye Kalkınma Bankası A. Ş.,* APM/28 (KİG-26), ss. 1–2.

Erdugan, F., Yörübulut, S., "Yatan Hastalar İçin Probit Regresyon Analizi İle Hastanenin Tekrar Tercih Edilebilirliğinin İncelenmesi", *Süleyman Demirel Üniversitesi Vizyoner Dergisi,* 2018/9, (22), ss. 13–20.

Ersöz, F., "OECD'ye Üye Ülkelerin Seçilmiş Sağlık Göstergelerinin Kümeleme ve Ayırma Analizi ile Karşılaştırılması", *Turkiye Klinikleri J Med Sci,* 2009/29, (6), ss. 1650–1659.

Girginer, N. ; Cankuş, B. "Tramvay Yolcu Memnuniyetinin Lojistik Regresyon Analiziyle Ölçülmesi: Estram Örneği", *Yönetim ve ekonomi,* 2008/15, (1), ss. 181–193.

Gujarati, D. N., (2005), Temel Ekonometri, çev. Ümit Şenesen, Gülay Günlük Şenesen, (1. baskı), Literatür Yayınları, İstanbul.

Gürsakal, N., (2015), Betimsel İstatistik, (8. baskı), Dora yayıncılık, Bursa.

Hosmer, D. W., Javanovic, B. and Lemeshow, S., "Best Subsets Logistic Regression", *Biometrics,* December 1989/45, (4), ss. 1265–1270.

Hosmer, D. W., and Lemeshow, S., (1989), Applied Logistic Regression, (First Edition), John Wiley & Sons, 307, New York-USA.

Işık ,S., Duman, K. ve Korkmaz A., "Türkiye Ekonomisinde Finansal Krizler: Bir Faktör Analizi Uygulaması", *Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi,* 2004/19 (1), ss. 45–69.

Johnson, W., "Influence Measures for Logistic Regression. Another Point of View", *Biometrica,* 1982/72, (1), ss. 59–65.

Kalaycı, Ş., (2010), SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri, (5. baskı), Asil yayıncılık, Ankara.

Karaağaoğlu, E., "Bilimsel Makalelerde İstatistiğin Kullanımı, Sunumu ve Sık Yapılan Hatalar", *Sağlık Bilimlerinde Süreli Yayıncılık-2005,* ss. 165–169.

Kaya, M., Genç, M., Kaya, B. ve Pehlivan, E., "Tıp Fakültesi ve Sağlık Yüksekokulu Öğrencilerinde Depresif Belirti Yaygınlığı, Stresle Başaçıkma Tarzları ve Etkileyen Faktörler", *Türkiye Psikiyatri Dergisi,* 2007/18, (2), ss. 137–146.

Kleinbaum, D. G., Kupper, L. L., Chambless, L. E., "Logistic Regression Analysis of Epidemiologic DATA: Theory and practice", *Communications in Statistics – Theory and Methods,* 1982/11, (5), ss. 485–547.

Kılıç S., "Bağıntı Analizi Sonuçlarının Yorumlanması", *Journal of Mood Disorders,* 2012/2, (4), ss. 191–193.

Kırcı, Çevik, N. ve Korkmaz O., "Türkiye'de Yaşam Doyumu Ve İş Doyumu Arasındaki İlişkinin İki Değişkenli Sıralı Probit Model Analizi", *Niğde Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,* 2014/7, (1), ss. 126–145.

Kitiş, Y., Bilgili, N., Hisar, F. ve Ayaz, S., (2009), "Yirmi Yaş ve Üzeri Kadınlarda Metabolik Sendrom Sıklığı ve Bunu Etkileyen Faktörler", *Avrupa Jinekoloji Derneği Kongresi,* Roma, İtalya, 2009.

Korkmaz, S., Göksülük, D. ve Zararsız, G., (2014), "MVN: Çok Değişkenli Normal Dağılıma Uygunluğun Belirlenmesi için Bir R Paketi", *XVI. Ulusal Biyoistatistik Kongresi*, Antalya, 2014.

Kutlar, A., (2007), Ekonometriye Giriş, (1. baskı), Nobel Yayınları, Ankara.

Mardia, K., V., "Measures of Multivariate Skewness and Kurtosis with Applications", *Biometrika,* 1970/57, (3), ss. 519–530.

Mecklin, C. J. and Mundfrom D. J., "A Monte Carlo Comparison of the Type I and Type II Error Rates of Tests of Multivariate Normality", *Journal of Statistical Computation and Simulation,* 2005/75, (2), ss. 93–107.

O'Neill, T. J. and Barry, S. C., "Truncated Logistic Regression", *Biometrics,* 1995/51, (2), ss. 533–541.

Oğuzlar, A., "Hanehalkı Tipi ve Kır-Kent Ayırımının Diskriminant Analizi ile İncelenmesi", *Akdeniz Üniversitesi İİBF Dergisi,* 2006/6, (11), ss. 70–84.

Öven, V. A. ve Pekdemir, D., "Faktör Analizi ile Ofis Kira Değerini Etkileyen Parametrelerin Belirlenmesi", *İTÜDERGİSİ/a Mimarlık, Planlama, Tasarım,* 2005/4, (2), ss. 3–13.

Özdamar, K., (1999), Paket Programlar ile İstatistiksel Veri Analizi, (2. baskı), Kaan Kitabevi, Eskişehir.

Öztürk, A. ve Türker, M. F., "Devlet Orman İşletmelerinin Gruplandırılmasında Çok Değişkenli İstatistiksel Analizlerin Kullanımı", *Artvin Çoruh Üniversitesi Orman Fakültesi Dergisi,* 2010/11, (2), ss. 20–29.

Prentice, R., "Use of the Logistic Model in Redrospective Studies", *Biometrics,* 1976/32, (3), ss. 599–606.

Pregibon, D., (1981), "Logistic Regression Diagnostics", *The Annals of Statistics,* 1981/9, (4), ss. 705–724.

Robert, G., Rao, N. K. and Kumar, S., (1987), "Logistic Regression Analysis of Sample Survey Data, *Biometrika,* March 1987/74, (1), ss. 1–12.

Sağlam, M, "Çok Değişkenli İstatistiksel Yöntemler ile Toprak Özelliklerinin Gruplandırılması", *Topraksu Dergisi*, 2013/2, (1), ss. 7–14.

Sayın, S., "Bilimsel Araştırmalarda Yapılan İstatistiksel ve Yöntembilimsel Hatalar–II: Grafik, Tablo ve Gösterim Hataları", *Türk Eğitim Bilimleri Dergisi*, Kış 2010/8, (1), ss. 117–143.

Sayın, S., "Bilimsel Araştırmalarda Yapılan Bazı İstatistiksel ve Yöntembilimsel Hatalar–III: Güvenirlik Kestirimlerine Yönelik Hatalar", *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi,* 2008, (15), ss. 53–69.

Selim, S. ve Sarıbay, E., "Yabancı Dil Eğitimi Veren Özel bir Eğitim Kurumundaki Öğrencilerin Beklentilerinin Araştırılması", *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 2003/5 (2), ss. 104–113.

Tabachnick, B. G., ve Fidell, L. S., (2015), Çok Değişkenli İstatistiklerin Kullanımı, çev. ed. Prof. Dr. Mustafa Baloğlu, (6. baskı), Nobel yayınları, Ankara.

Tarı, R., (2008), Ekonometri, (5. baskı), Kocaeli Üniversitesi Yayınları, İstanbul.

Tatlıdil H., (2002), Uygulamalı Çok Değişkenli İstatistiksel Analiz, Akademi Matbaası, Ankara.

Tavşancıl, E., (2014), Tutumların Ölçülmesi ve SPSS ile Veri Analizi. (4. baskı), Nobel Akademik Yayıncılık, Ankara.

Tektaş, N., (2014), "Classification Performance Comparison of Discriminant Analysis and Logistic Regression Analysis", *Turkish Studies – International Periodical For The Languages, Literature and History of Turkish or Turkic*, Winter 2014/9, (2), ss. 1517–1527.

Toktay, Y., (2017), *Çok Değişkenli İstatistik Analiz Yöntemler: Faktör Analizi Ve Diskriminant Analizinin Iğdır Üniversitesi Öğrencileri Üzerine Uygulaması,* (Yüksek Lisans Tezi), Iğdır Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni Anabilim Dalı, Iğdır.

Topak, M. S., "İmalat Sanayinde Firma Risklerinin Belirlenmesi: Kümeleme Analizi Yöntemiyle Ampirik Bir Çalışma", *İstanbul Üniversitesi İktisat Fakültesi Ekonometri ve İstatistik Dergisi*, 2010, (11), ss. 100–127.

Turanlı, M., Güriş, S., (2015), Temel İstatistik, (5. baskı), Der yayınları, İstanbul.

Uysal, F. N., Ersöz, T. ve Ersöz, F., "Türkiye'deki İllerin Yaşam Endeksinin Çok Değişkenli İstatistik Yöntemlerle İncelenmesi", *Ekonomi Bilimleri Dergisi*, 2017/9, (1), ss. 49–65.

Ünlükaplan, Y., (2008), *Çok Değişkenli İstatistiksel Yöntemlerin Peyzaj Ekolojisi Araştırmalarında Kullanımı,* (Doktora Tezi), Çukurova Üniversitesi, Adana.

Yaşlıoğlu, M. M., "Sosyal Bilimlerde Faktör Analizi ve Geçerlilik: Keşfedici ve Doğrulayıcı Faktör Analizlerinin Kullanılması", *İstanbul Üniversitesi İşletme Fakültesi Dergisi,* Special Issue/Özel Sayı 2017/46, ss. 74–85.

Yücel Toy, B., Güneri Tosunoğlu N., "Sosyal Bilimler Alanındaki Araştırmalarda Bilimsel Araştırma Süreci, İstatistiksel Teknikler Ve Süreci, İstatistiksel Teknikler Ve Yapılan Hatalar", *Ticaret ve Turizm Eğitim Fakültesi Dergisi,* 2007, (1), ss. 1–20.

Yıldız, Ö., (2011), *Serebral Palsili Çocuklarda Konvülsiyonu Etkileyen Risk Faktörlerinin Logistik Regresyon İle İncelenmesi*, (Yüksek Lisans Tezi), Yüzüncü Yıl Üniversitesi Sağlık Bilimleri Enstitüsü, Van.

Türkçe İstatistik Rehberi
<http://www.istatistik.gen.tr/?p=100 > (Erişim tarihi: 24.10.2018)

ankara.edu.tr (Pdf dosyası)
<https://acikders.ankara.edu.tr/pluginfile.php/1382/mod_resource/content/2/B9_Normal%20Da%C4%9F%C4%B1l%C4%B1m.pdf > (Erişim tarihi: 25.09.2019)

Kemal Doymuş
<https://kemaldoymus.files.wordpress.com › 2009/12 › ders_3 > (Erişim tarihi: 25.09.2019)

anadolu.edu.tr (Pdf dosyası)
<http://eczacilik.anadolu.edu.tr/bolumSayfalari/belgeler/ecz2014%2012_20140527094539.pdf > (Erişim tarihi: 25.09.2019)