Halil Yurdugül
yurdugul@hacettepe.edu.tr

Bilimsel Araştırmalarda İstatistiksel Yöntemler

Doğa ve davranış bilimlerine konu olan olgu ve olaylar belli bir "sistem" çerçevesinde gelişirler. Bu olgu ve olaylara göre davranış gösteren sistemin özelliklerini ortaya koymak bilimsel çalışmaların temelini oluşturur.

Sistem, genel yapısı itibari ile üç ayrı kategoride ele alınabilir. Bunlar sırasıyla; ulaşılamayan, ulaşılabilen ya da ulaşılabilir olmasına rağmen üzerinde çalışması güçlüklere yol açan sistemler. Dolaysız olarak sistem üzerinde çalışmanın en büyük avantajı, elde edilen bulguların kesinlik taşımasıdır. Bu şekilde elde edilen bulguların genellenme ihtiyacı söz konusu değildir.

Ancak sistem, bazen ulaşılmayabilir ya da ulaşılması zaman ve maliyet açısından sıkıntı yaratacak büyüklükte olabilir. Bu durumda, sistemi en iyi temsil edebilecek model ya da maket üzerinde elde edilen bulgular sayesinde sistemin özellikleri kestirilir.

Model ve maket arasındaki en önemli fark, maketlerin ayrıntılara fazla yer vermemesidir. Bu nedenle bilimsel araştırmalarda maket üzerinde çalışma tercih edilmemekle birlikte maket yöntemi genellikle teknolojik araştırmalar da sıkça kullanılmaktadır.

Bilimsel Araştırmanın Temelleri

Bilimin günümüze kadar yapılmış pek çok tanımı vardır. Bunlarda birinde bilim, doğa ve toplum arasındaki ilişkileri ortaya koymaya yönelik etkinlikler bütünü olarak tanımlanır. Bu bakış açısı, materyalist felsefenin diyalektik açılımına da karşılık gelmektedir. Bu açılıma göre; doğadaki bütün nesneler neden-sonuç ilişkisi (nedensel ilişki) içindedir ve bu ilişki içersinde niceliksel-niteliksel bir değişim (birlikte değişim-covariate) gösterirler.

Bu yaklaşımı Yıldırım ve Şimşek pozitivist/akılcı paradigmanın özellikleri olarak belirtir (Şimşek ve Yıldırım; s:3). Günümüze kadar bir değişim gösteren bilimsel paradigmanın diğer özellikleri değişirken bu iki özellik, doğanın temel özelliğini yansıttığı için değişmeyen özellik olarak kaldılar.

Dr. Şenocak ise bu iki özelliği farklı bir isimlendirme ve açıklama yoluna gitmiştir.[1]

> Evrendeki "soruların" yanıtlarının aranması yöntemlerinden en geçerli ve gerçekçisi olan "nedensellik" soruşturması kaba çizgilerle iki biçimde ortaya çıkabilir;
> **- Nedensel bağ ("Causal" bağ):**
> Varlığı öncül olan bir olayın, daha sonra ortaya çıkan bir olayın varlığının nedeni olması durumudur. Gerçekte doğada, bir sonucun (bağımlı öge) katı ve gerekirci bir yaklaşımla sadece tek bir nedene bağlı olmasına hele de

[1] http://abone.superonline.com/~senocakbiyo/ktp.htm

tamamen kestirilebilir bir bağıntı ile bağlı olmasına hemen hiç rastlanmaz. Nedensel bağlarda; çoklu öncül ögeler, sonuç çeşitliliği ve sonuç belirme eşiğinin değişkenliği gibi özel durumlar kaçınılmaz olarak vardır ve çözümlemelerde de bunlar kesinlikle göz önünde bulundurulur. Yargılamalar da bu çerçevede ancak "olasılığa" bağlı düzeylerle belirtilebilir .Örneğin sigara, akciğer kanseri oluşumuna (sonuç), temel bir nedenmiş gibi görünmektir, ancak sonucu tek başına ve tümü ile sigaraya bağlamak her halde söz konusu değildir, (Yaş, cinsiyet, vs. sonucu etkileyecektir) çözümleme, bu tür ikincil etkenleri de göz önünde bulundurarak hangi olasılık sınırları çerçevesinde sigara içimi ile kanserin varlığının bağdaşabileceğini ortaya koyabilir.

- Birlikte değişim bağı ("Covariational" bağ):
İki ayrı olgudan birinin neden, diğerinin sonuç olma özelliğinin kesin belirlenemediği veya zaten bununla ilgilenilmediği, ancak bu iki değişkendeki ölçümsel farklılaşmaların birbirlerinden etkilendiği - birinin artmasının diğerinde de bir artış veya azalma oluşturması - olaylarla karşımıza çıkar. Bu değişim birlikteliği, iki olayın büyüklük düzeyleri arasında kabaca da olsa matematiksel bir fonksiyon (regresyon) ortaya koyacak kadar belirgin ise, elimizde, değişkenlerden birinin değerine bakıp, diğerini büyük bir olasılıkla "kestirebilme" gibi bir olanak bulunur.

Şenocak'ında belirttiği gibi, olaylar ya da olgular arasındaki ilişkinin irdelenmesi ister nedensellik ilişkisi isterse de birlikte değişimi bazında olsun, uygun bir araştırma yöntemi ile gerçekleştirilebilir.

Bu araştırmalara yönelik en genel sınıflama nicel ve nitel araştırma yöntemleri olarak bilinmektedir. Bu çalışmada nicel araştırma yöntemleri ele alınacaktır.

Nicel araştırmalara ilişkin bir çok genellemeler vardır. Bunların en genel biçimi aşağıda verilmiştir.[2]

1) Analitik Çözümleme
2) İteratif Çözümleme
3) Simülatif Çözümleme

Analitik Çözümleme

Çözüm sonuçları,mutlak doğru olarak gösterilir/kullanılır. Genellikle hata terimi içermeyen (deterministik) matematiksel çözümleri içerir. Analitik çözümleme

$$(1/3)+(1/3)+(1/3) \neq 1$$

eşitsizliğini kabul ederek sonucu

$$(1/3)+(1/3)+(1/3) = 0,99999......$$

olarak ele alır.

[2] H.Ü. İstatistik Bölümü, Benzeşim (simulation) ders notlarından. 1996

Analitik çözümleme, genellikle sistemde yapılan çalışmalarda kullanılır. Bu nedenle elde edilen bulgular dolaysız olarak, kesin bilgi niteliği taşımaktadır.

İteratif Çözümleme

Analitik çözümlemede belirtilen kesin çözüme, belli bir yaklaşıklıkla (approach) ulaşılır. Bir diğer ifade ile kesin bilgiye belli bir ε komşuluğu ile yakınlaşır. Bu durum genellikle olasılıksal süreçlerde (probabilistic) kullanılır. İstatistik ve sayısal çözümleme (numerical analysis) bilimlerinin temelini oluşturur.

İteratif yöntemler, genellikle model üzerinde yapılan çalışmalarda kullanılır. Modelden elde edilen bilgiler, ε komşuluğunda sistem hakkında bilgi veir.

Simülatif Çözümleme

Benzeşim (simulation) olarakta adlandırılan bu yöntem, diğerlerine göre daha az duyarlıdır. Genellikle sisteme ulaşılamadığı durumlarda ya da sistemin tanımlanamadığı durumlarda kullanılır. Sisteme ulaşılamadığı için, modelden elde edilen sonuçların asimtotik davranışının, sistemin özelliklerini yansıttığı vurgulanır.

Bu yöntemin bir diğer özelliği ise; model parametrelerinin kontrol altına alınarak parametrik davranışlarındaki değişimlerin değişik ölçütler altında gözlenebilmesini olanaklı hale getirmektedir.

İstatistiksel Çözümlemeler

İstatistiksel çözümlemeler, değişkenlerin sahip oldukları fonksiyonların bağıntılarının irdelenmesine yöneliktir. Fonksiyonel yaklaşımlar ya da evren-örneklem bağıntısını açıklamaya yönelik istatistiksel modeller, değişkenlerin yapısına, sayısına ve kesitine göre değişiklikler gösterir.

Değişken Yapısı:

Veriler ya da verileri temsil eden değişkenler kesikli veya sürekli bir yapıda olabilir. Bu ayırımın temel ölçütü ise değişkenlerin sahip oldukları tanım kümesidir. Değişkenlerin alabileceği değerlerin oluşturduğu tanım kümesi tam değerlerden oluşturduğu tanım kümesi tam değerlerden oluşuyorsa bu tür değişkenlere kesikli kesikli değişkenler adı verilir. Ancak, değişkenlerin tanım kümesi, aynı zamanda bir tanım aralığı ise sürekli değişkenler olarak adlandırılır. Değişkenlerin yapısı, kurulacak istatistiksel modellerde tamamen bir değişiklik gösterir.

Değişken Sayısı:

İstatistiksel çözümlemelere konu olan araştırmalardaki değişken sayısı, kurulacak olan istatistiksel modelleri tamamen etkilemektedir. Yalnızca bir değişkenden sözediliyorsa tekdeğişkenli (univariate), değişken sayısı, en az iki olduğunda ise çokdeğişkenli (multivariate) çözümleme yöntemleri üzerinde çalışmak söz konusudur. Bazı kaynaklarda ise bu gruplamanın içinde ikideğişkenli (bivariate) durumlar ayrı ele alınmaktadır.

Tekdeğişkenli bir süreç ya da durum üzerinde çalışılıyorsa, varolan durumu/süreci tanımak için betimsel istatistiklere ihtiyaç vardır. Eğer sistem birden fazla değişken içeriyorsa çokdeğişkenli istatistiksel çözümlemelere başvurulur.

Değişkenlerin Kesiti

Verilerin elde edildiği zaman kesiti, istatistiksel çözümlemelerin farklılaşmasına neden olmaktadır. Veriler, zamanın belirli bir kesitinde elde edilmiş ise bu tür verilere yatay-kesit veriler, zamanın değişik kesitlerinde elde edilmiş ise bu tür verilere de dikey-kesit veriler adı verilemktedir.

Yatay-kesit verilerin en büyük özelliği verilerin homojenlik göstermeleridir. Bu tür çalışmalarda homojenlik, deneyin yapıldığı ortamın özelliklerinin aynı özellikleri taşıması anlamındadır.

Dikey-kesit verilerin en çok kullanıldığı yer, zaman serileri olarak adlandırılan ve zamanın değişken olarak ele alındığı uyguamalardır. Bu tür zamanın değişken olarak ele alındığı çözümlemelerde, ya zaman değişkeninin deney düzenine etkisi ya da zamanın değişik düzeylerindeki bağımlı değişkendenin değeri kestirilir (forecast, prediction).

İstatistiksel Modeller ve Çözümlemeler

Çok değişkenin söz konusu olduğu durumlarda bir modelden söz etmek gereklidir. İstatistiksel modeller, yukarıda özetlenen, ikideğişkenli ve çokdeğişkenli olarak işleyen sistemlerde, değişkenlerin yapısına, sayısına ve kesitine göre değişiklikler gösterirler.

<table>
<tr><td colspan="2" rowspan="2"></td><th colspan="2">BAĞIMLI</th></tr>
<tr><th>Kesikli</th><th>Sürekli</th></tr>
<tr><th rowspan="2">BAĞIMSIZ</th><th>Kesikli</th><td>* Ki-Kare Çözümlemesi
* Log-Linear Çözümleme
* Uygunluk Çözümlemesi (Corresponds Analysis)</td><td>* Varyans Çözümlemesi
* Kovaryans Çözümlemesi
* Regresyon Çözümlemesi (Dummy Değişkenli)</td></tr>
<tr><th>Sürekli</th><td>* Lojistik Regresyon
* Diskriminant Çözümlemesi</td><td>* Regresyon Çözümlemesi
* Korelasyon Çözümlemesi</td></tr>
</table>

Yukarıda da özetlendiği gibi, model yapıları temelde bağımlı ve bağımsız değişkenlerin yapısına göre değişiklik göstermektedir. Ancak bu değişkenler, yatay-kesit veriler üzerine kurulu ve istatistiksel modellerde, homojen ortamlarda kurulan istatistiksel modellerdir.

Modeller arasındaki bir diğer ayırım ise; değişkenler arasındaki "nedensellik bağ" ya da "birlikte değişim bağ" ının mı irdeleneceğidir. Bir değer ifade ile, bağımsız değişkenlerin bağımlı değişken üzerindeki etkisi mi yoksa ilişkisi mi araştırma konusudur?

Ki-kare çözümlemesi

Ki-kare çözümlemesi, bağımlı ve bağımsız değişkenlerin kesikli olduğu durumlarda kullanılır. Araştırma konusu, daha çok iki değişkenin arasındaki ilişki ya da bağımsızlık olduğunda geçerlidir. Ki-kare çözümlemesi için olumsallık çizelgelerinden (contingency table) yararlanılır.

Ki-Kare=(Gözlenen-Beklenen)^2 / Beklenen^2

Elde edilen bu değer, tablo değeri ile karşılaştırılır ve "Değişkenler arasında ilişki yoktur" hipotezi test edilir. Ancak bu yargıya ek olarak, değişkenler arasındaki ilişkinin miktarınıda belirlemek olanaklıdır. Bu amaçla ortaya çıkarılan bir çok ilşki ölçüsü (measurements of association) vardır (Yurdugül, 1997).

Log-Linear Çözümlemesi

Ki-kare çözümlemesinin daha genel, çok değişkenli halidir. Bir bakıma varyans çözümlemesinin kesikli biçimi olarakta görülebilir.

	1	**j**	**Toplam**
1	π_{11}	π_{1j}	π_{1+}
i	π_{i1}	π_{ij}	π_{i+}
Toplam	π_{+1}	π_{+j}	N

Bilindiği gibi olumsalık tablolarının bileşik olasılıkları içeririr (π_{ij}). Ancak satır ve sütunlarda bileşen olasılıklardan oluşur (π_{+j} ve π_{i+}). Bağımsızlık koşulu altında beklenen birleşik olasılıklar aşağıdaki gibidir (Agresti, A: sy:25).

$\pi_{ij}=(\pi_{+j})(\pi_{i+})$

$m_{ij}=N.\ (\pi_{+j})(\pi_{i+})$

$\log(m_{ij})=\log N+\log(\pi_{i+})+\log(\pi_{+j})$

Buradaki eşitlikte, varyans çözümlemesinde olduğu gibi

$\mu_{ij}=\log(m_{ij})$

$\log(m_{ij})=\mu+\lambda_i^{x}+\lambda_j^{y}+\lambda_{ij}^{xy}$

$\lambda_i^x=\mu_{i+}-\mu$ (x değişkeninin etkisi)

$\lambda_i^y=\mu_{+j}-\mu$ (y değişkeninin etkisi)

$\lambda_{ij}^{xy}=\mu_{ij}-\mu_{i+}-\mu_{+j}-\mu$ (Kesişim-Interaction- terimi)

Görüldüğü gibi log-linear modelin, varyans çözümleme modelinin kesikli verilere uyarlanmış biçim olduğunu söylemek yanlış olmayacaktır.

Uygunluk Çözümlemesi (Correspond Analysis)

Uygun Getirme Çözümlemesi olarakta adlandırılan bu model, ki-kare çözümlemesinin görsel ağırlıklı bir şeklidir. Bir diğer ifade ile genellenmiş ki-kare olarakta ifade edilmektedir.

Bilindiği gibi klasik ki-kare çözümleme modeli iki değişken ile kurulabilmektedir. Log-linear model ve uygunluk çözümleme modelleri daha üst çok değişkenle çalışma olanağı sağlamaktadır. Log-linear model, daha çok gözlenen ile beklenen değer arasındaki değişimin bileşenlerini ele alırken, uygunluk çözümlemesi daha karmaşık bir kuramsal temele dayanmaktadır.

Varyans Çözümlemesi (ANOVA)

Log-linear modelin sürekli değişkenler için kullanılan biçimidir. Bağımlı değişken üzerindeki bağımsız değişkenin (faktör değişkeni) değişik düzeylerindeki etkilerini ortaya çıkarmakta kullanılır. Bağımsız değişkenin herbir düzeyindeki varyansların homojen olduğunda geçerlidir. Bir diğer ifade ile, homojen bir deney düzeninde yapılan uygulamalar uygulanır.

Varyans çözümlemesine ilişkin tek etkenli doğrusal model aşağıdaki gibidir:

$Y_{ij}=\mu+\varepsilon_{ij}$

Ancak bağımsız değişkenin çeşitli düzeylerinin etkisi aynı olmadığı için modele bunu da eklemek gereklidir. Tek etkenli doğrusal model aşağıdaki gibidir.

$Y_{ij}=\mu+\tau_j+\beta_i+\varepsilon_{ij}$

τ_j: Etkenin j. Düzeyini göstermektedir.

Bu doğrusal modelin açılımı aşağıdaki gibidir.

$(Y_{ij}-\mu)^2=(Y_{ij}-\mu_j)^2+(\mu_j-\mu)^2$

GenelKT=GİKT+GAKT

GİKT=Gruplar İçi Kareler Toplamı

GAKT=Gruplar Arası Kareler Toplamı

Buradaki sonuç bir diğer ifade ile, verilerin genel varyansın bileşenlerine ayırarak bir F değeri etmektir. Bu amaçla elde edilen değerler ANOVA (Analysis of Variance) çizelgelerinde gösterilir (İnal C ve Günay S. Sy:480).

Değişim Kaynağı	S.D.	Kareler Toplamı	Kareler Ortalaması	F
Gruplararası (etken)	k-1	$n\Sigma(\mu_j-\mu)^2$	K.T./S.D.	GAKO/ GİKO
Gruplariçi (Hata)	k(n-1)	$\Sigma(y_{ij}-\mu_j)^2$	K.T./S.D.	
Toplam	(kn-1)	$\Sigma(y_{ij}-\mu)^2$		

Burada F testine ilişkin yoklanacak olan hipotez:

$H_0=\mu_1+\mu_2+\mu_3+...+\mu_k=\mu$

Burada k; etkenin düzey sayısını göstermektedir.

Eğer etken sayısı (bağımsız değişken) 2 olsaydı bu sefer doğrusal model aşağıdaki gibi olacaktı.

$Y_{ij}=\mu+\tau_j+\beta_i+\varepsilon_{ij}$

τ_j: 1. etkenin j. Düzeyini göstermektedir.

β_i: 2. etkenin i. Düzeyini göstermektedir

İki etkenli ya da daha fazla etkenli deney düzenlerinde iki etkenin birbirleri ile olan etkileşimleri de (interaction) modele katılaak araştırılabilir. Böyle bir toplamsal model (addional model) aşağıdaki gibidir:

$Y_{ij}=A_i+B_j+AB_{ij}+e_{ij}$

A_i: A etkeninin i. düzeyi
B_j: B etkeninin j. Düzeyi
AB_{ij}: Ave B etkenlerinin kesişimi

Tek Etkenli Rastgele Blok Düzeni:

Deney düzenindeki homojenlik varyans çözümlemesi için zorunlu bir koşuldur. Bunun bir diğer ifadesi ise tam rastgeleliktir. Ancak bazen deneyin yapıldığı ortamların etkisi de inelenmek üzere modele dahil edilebilir. Bunu faktör değişken ile aynı tutulmaması gerekmektedir.

Halil Yurdugül
yurdugul@hacettepe.edu.tr

Örneğin A ve B gibi iki öğretim yönteminin öğrencinin performans artışındaki etkisi araştırılmak üzere bir deney düzeni oluşturulsun. Deney için iki ayrı sınıfta A ve B yöntemleri uygulanıp öğrencinin başarısı araştırıldığında ortaya tek etkenli bir deney düzeni çıkar.

Ancak öğrenciler sınıflara rastgele dağıtılmadığı için sınıftaki öğrencilerin genel başarısının homojenliği bozup bozmadığının irdelenmesi gerekmektedir. Ya da öğretmenlerin performansının eşit olduğu kabul edilir. Ancak öğretmenlerin performansı da modele dahil edilebilir. Bu durumda öğrenmenler birer blok olarak modele dahil edilebilir ve rastgelelik üzerindeki etkisi ayrıştırılabilir.

Bu deney düzenine ilişkin doğrusal model aşağıdaki gibidir:

$Y_{ij}=\mu+\tau_j+\beta_i+e_{ij}$

β_i: i. blok etkisi

Böyle bir deney düzeninden oluşan ANOVA çizelgesi aşağıdaki gibidir (John, P: sy:58).

Değişim Kaynağı	S.D.	KT	F
Etken	k-1	$b\sum(\bar{y}_{i.}-\bar{y}_{..})^2$	KT/S.D.
Blok	b-1	$k\sum(\bar{y}_{.j}-\bar{y}_{..})^2$	KT/S.D.
Hata	(k-1)(b-1)	$\sum(\bar{y}-\bar{y}_{.j}-\bar{y}_{i.}-\bar{y}_{..})^2$	
Genel	N-1		

Kovaryans Çözümlemesi (ANCOVA)

Kovaryans çözümlemesi, varyans çözümlemesindeki homojen olmama durumunu ortadan kaldırmak için kullanılır. John, P, (John, P: sy:60), bunu rastgele blok düzeninin bir başka biçimi olarakta ifade etmektedir. ANCOVA aynı zamanda öntest-sontest deney düzeni olarakta adlandırılmaktadır.[3]

A ve B gibi iki öğretim programı sözkonusu olduğunda, bu programların öğrenci başarısı üzerine olan etkisi tek etkenli bir deney düzeni olarak ele alınabilir.

Ancak öğrencilerin ilgili üniteye ilişkin hazır bulunuşlukl davranışları (X) ve süreç sonu başarı puanları (Y) olduğuna göre:

[3] http://trochim.human.cornell.edu/kb/expcov.htm

Halil Yurdugül
yurdugul@hacettepe.edu.tr

A Öğretim Yöntemi		B Öğretim Yöntemi	
I. Gruptaki Öğrencilerin		II. Gruptaki Öğrencilerin	
Hazırbulunuşluk (X)	Başarı Puanı (Y)	Hazırbulunuşluk (X)	Başarı Puanı (Y)
35	60	58	79
57	81	39	55
:	:	:	:

Burada kurulacak doğrusal model aşağıdaki gibidir:

$Yij=\mu+\tau_j+\beta X_{ij}+e_{ij}$

Bu modelin bileşenlerini göstermek üzere oluşturulacak ANCOVA (**An**alysis of **Cova**riance) çizelgesi aşağıdaki gibidir.

ANCOVA Çizelgesi

Kaynak	**S.D.**	**x**	**xy**	**y**
Model	N	$\Sigma\Sigma x_{ij}^2$	$\Sigma\Sigma x_{ij}y_{ij}$	$\Sigma\Sigma y_{ij}^2$
Ortalama	1	$X^2_{..}/N$	$X_{..}Y_{..}/N$	$Y^2_{..}/N$
Etken	k	T_{xx}	T_{xy}	T_{yy}
Hata	N-k	E_{xx}	E_{xy}	E_{yy}
Genel	N-1	S_{xx}	S_{xy}	S_{yy}

Ancova Çizelgesi ile modele ilişkin bileşeler ortaya çıkarılır. Ancak nihai test için buradan varyans çözümlemesine (ANOVA) geçilir.

Kaynak	**S.D.**	**KT**
Model Uyumu	1	$(S_{xy})^2/S_{xx}$
Etken	k-1	$(S_{yy}-(S_{xy})^2/S_{xx})-(E_{yy}-(E_{xy})^2/E_{xx})$
Hata	N-k	$(E_{yy}-(E_{xy})^2/E_{xx})$
Genel	N-k-1	S_{yy}

Buradaki bileşenlerin hesaplanması için gerekli eşitlikler EK 1'de verilmiştir.

Halil Yurdugül
yurdugul@hacettepe.edu.tr

Diskriminant Çözümlemesi

Diskriminant Çözümlemesi, diğerlerinde olduğu gibi dolaysız bir modelleme olmayıp, özünde bir sınıflama yapısına sahiptir. Örneğin belirli bir ölçüte göre verilerin sınıflanması gerektiğinde kullanılır. Ancak bir sonraki konu olan lojistik fonksiyonla en önemli farkı, polythmous (ikiden fazla kaegorisi olan) değişkenler için uygun olmasıdır.

Diskriminant çözümlemesinin bir başka önemli özelliği de yanlış sınıflanmış verileri de ortaya çıkarmasıdır.

Bu konuda sınıflama için gerekli olan ölçüt, diskriminant fonksiyonudur. Diskriminant fonksiyonu aşağıdaki bağıntı üzerine kuruludur.

$y_i=a_3x_{i3}+ a_3x_{i3}+ a_3x_{i3}+....+ a_px_{ip}$

xi: orijinal değişkenler.

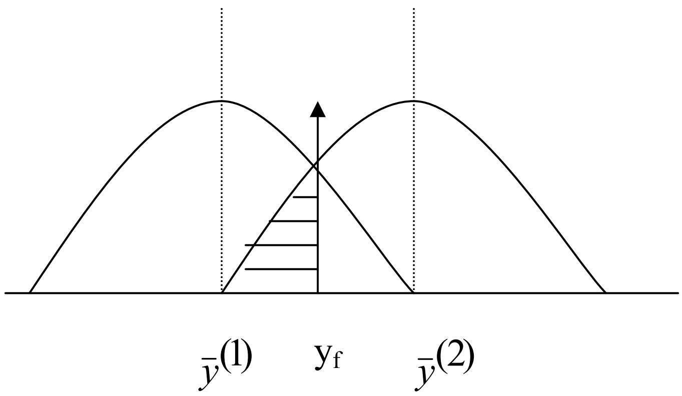
ai: bu değişkenlere ilişkin ağırlıklar.

Böyle bir fonksiyon bulunurken gruplararası varyansın enbüyüklenmesi gerekir (Tatlıdil, H. Sy:202).

F=max {Gruplararası Varyans/Gruplariçi Vaaryans}

Bağımlı değişken iki kategoriden oluştuğu düşünülsün. Yani 0 ve 1 değerleri alsın (evet-hayır, doğru-yanlış vs)

Şekil 1: Diskriminant fonksiyonunun ayırıcılığı.



Bir başarı tetinde doğru ya da yanlışları ele aldığımızda, doğruların ya da yanlışların kendi içlerinde bir dağılımı olacaktır. Bu dağılımların ortalaması aşağıdaki gibidir.

$$\bar{y}^{(1)} = a_1\bar{x}_1^{(1)} + a_2\bar{x}_2^{(1)} + \ldots. + a_p\bar{x}_p^{(1)}$$
$$\bar{y}^{(2)} = a_1\bar{x}_1^{(2)} + a_2\bar{x}_2^{(2)} + \ldots. + a_p\bar{x}_p^{(2)}$$

İki ayrı ortalama olmasına rağmen varyans olarak, ortak bir varyanstan sözetmek gereklidir. Ortak varyans aşağıdaki gibi gösterilebilir (Tatlıdil, H. Sy:204).

$$Var(y) = \bar{y}^{(1)} - \bar{y}^{(2)}$$

Şekil 1'de gösterilen yf ayırım fonksiyonunun sayısal eşitliği ya da doğru/yanlış sınıflama aşağıda verilmiştir.

$$y_f = \frac{\bar{y}^{(1)} + \bar{y}^{(2)}}{2}$$

Lojistik Regresyon

Özellikle sağlık bilimlerinde sıkça kullanılan bu çözümleme türü aynı zamanda davranış bilimlerinde de kullanım alanları çoktur. Diskriminant çözümlemesinde genellikle bir sınıflama yapılmaktadır. Lojistik regresyonda da aynı durum söz konusudur. Özellikle iki sınıflı kategorik (Dichothomous) veriler için önerilmektedir.

Eğitim alanında özellikle Örtük Özelliker Kuramı'nda (Latent Trait Theory) alanında ele alınmaktadır. Bağımsız değişken olarak, test başarı puanı, bağımlı değişken olarakta (0 ve 1) madde puanı olarak modellenmektedir.

Binary regresyon olarakta ifade edilebilen bu çözümleme türü, başarı üzerine kurulan yoğunluk fonksiyonu işlenmektedir (Andersen, E. Sy:354):

$$P(Y_i = 1) = \frac{e^{(\alpha + \beta X_j)}}{1 + e^{(\alpha + \beta X_j)}}$$

Regresyon Çözümlemesi

Regresyeon çözümlemesi iki anlamda ele alınmalıdır. Bunlardan ilki, değişkenler arasında fonksiyonel bir bağıntının irdelenmesi, diğeri ise bağıntının fonksiyonel olarak elde edilmesidir.

Evrene ilişkin bağıntı:

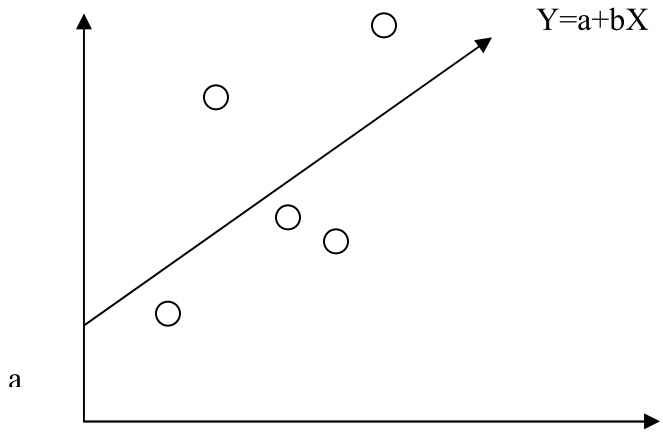
Y=α+βX+ε

Örnekleme ilişkin bağıntı:

Y=a+bX+e

Örneklemden elde edilen kestirim modeli:

Y=a+bX



Y=a+bX doğrusal bağlantısının kestirilmesi, katsayıların kestirilmesi ile ortaya çıkmaktadır.

$$b = \frac{\sum (x_j - \bar{x})(y_j - \bar{y})}{\sum (x_j - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

Burada a katsayısı denklem gereği olan bir katsayıdır ve X'in 0 değeri alması durumunda Y'nin alacağı değeri belirtir. Ancak b değeri, matematiksel olarak "eğim"e karşılık gelmektedir ve X'teki 1 birimlik artışa karşılık Y'deki değişimi ifade etmektedir.

Model kestirildikten sonra, 2. aşama da modelin geçerliliği test edilir. Burada β=0 yani X'teki artışın Y üzerinde önemli olduğunun bir araştırmasıdır. Bu nedenle modele ilişkin F testine ve ANOVA çizelgesine ihtiyaç vardır.

Kaynak	**S.D.**	**K.T.**	**K.O.**	**F**
Regresyon	k	$\sum(\hat{Y} - \bar{Y})^2$	K.T./S.D.	$\frac{KO_{Reg}}{KO_{Artık}}$
Artık	n-k-1	$\sum(Y - \hat{Y})^2$	K.T./S.D.	
Genel	n-1	$\sum(Y - \bar{Y})^2$		

Burada test edilecek hipotez H_0: β=0'dır. Eğer F değeri anlamlı bulunursa Evrende böyle bir doğrusal bağıntının var olduğu belirtilir

Eğer modelde bir tek bağımlı değişken ve yalnızca bir bağımsız değişken var ise bu tür modellere basit doğrusal regresyon modelleri adı verilir. Ancak bağımsız değişken sayısı birden fazla ise bu durumda üzerinde çalışılacak olan model, çoklu regresyon modelleridir.

$Y=\beta_0+\beta_1 X_1+\beta_2 X_2 +...+\varepsilon$

Ancak bu modele ilişkinin bağıntının anlamlı olabilmesi için iki önemli varsayım vardır. Bunlardan ilki hata teriminin $\varepsilon\sim(0,\sigma_\varepsilon)$ sabit varyanslı birer normal dağılım gösteriyor olmasıdır. Diğeri ise Xi'ler arasındaki covaryansın 0 olması bir diğer ifade ile bağımsız değişkenlerin kendi aralarında bağımsız olmasıdır.

1-) $\varepsilon\sim(0,\sigma_\varepsilon)$

2-) $Cov(X_i,X_j)=0$

Eğer bağımsız değişkenler arasında bir korelasyon var ise bunu gidermenin değişik yöntemleri vardır. Bunlardan ilki, temel bileşenler çözümlemesidir. Temel bileşenler çözümlemesi ilişkili değişkenler bir arada sınıflandırılarak yeni değişkenler olarak ifade edilirler ve birbirinden bağımsız ve yeni değişkenlere modelde yer verilir.

Bu yöntem zamanla, çokdeğişkenli çözümlemelerin bir çok konusunu oluşturmaktadır.

Korelasyon Çözümlemesi

Kanonik çözümlemesi, temel bileşenler çözümlemesi, faktör çözümlemesi, kümeleme çözümlemesi ve diğer yöntemlerin temel amacı ilişkili değişkenleri kümeleyerek boyut indirgeme ve ilişkili yeni değişkenler türetmek amacı taşımaktadır. Ancak bu çalışmada kanonik korelasyonçözümlemesi ele alınacaktır.

Kanonik Korelasyon Çözümlemesi

Bilindiği gibi korelasyon katsayısı ancak iki değişken arasından elde edilir. Değişken sayısı arttığı zaman diğer değişkenler sabit iken ancak iki değişken arasında korelasyon hesaplamak olanaklıdır.

ikiden fazla değişken arasında korelasyon üzerine son günlerde yapılan çalışmalar vardır. Bunlardan en önemlisi Harry Joe'nin bilgi kuramsal ilişki ve bağımlılık ölçüleridir (Yurdugül, H).

Ancak geçmişte sıkça kullanılan kanonik korelasyon çözümlemesi, yeni kanonik değişkenlere ulaşmakta ve orijinal değişkenler ile kanonik değişkenler arasındaki korelasyonları elde etmekte olanaklıdır.

En gelişmiş ve en karmaşık ilişki çözümlemesi olan kanonik korelasyon çözümlemesi, çok boyutlu kitleden çekilmiş iki ya da daha fazla değişken kümesi arasındaki ilişki ile ilgilenir. Raslantı değişkenler kümesinin doğrusal fonksiyonları arasındaki maksimum korelasyonları bulmaya çalışır (Tatlıdil, H. Sy:74).

X raslantı değişkeni kümesi (vektörü) ve Σ varyans-kovaryans matrisini ele aldığımızda;

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

x1 ve x2 raslantı vektörlerinin keyfi doğrusal bileşimleri u ve v olsun.

$$u = \alpha' x_1 \qquad v = \gamma' x_2$$

$u=a_1x_{11}+a_2x_{12}+...+a_px_{1p}$

$v=a_1x_{21}+a_2x_{22}+...+a_px_{2p}$

Faktör çözümlemesinde olduğu gibi öncelikle ağırlıklar ve yeni değişkenler elde edilmek suretiyle α ve γ katsayıları kestirilir.

Burada amaç yeni u ve v değişkenleri arasındaki korelasyonun maksimum olamasının sağlanmasıdır. Bu gerçekleştiğinde kanonik değişkenler ile orijinal değişkenler arasındaki korelasyona bakılır.

α ve γ katsayıları tek başına yorumlanabilineceği gibi Kor(u_i,x_j) değerleri arasındaki nicelik olarak en büyük değerlerde kendi içerisinde yorumlanabilir, değişkenler arasındaki önem derecesi belirlenebilir.

Değinilen Kaynaklar

Agresti Alan	Analysis of Ordinal Categorical Data, Newyork, 1984
Andersen, E. 1980	Discrete Statistical Models with Social Science Applications. Newyork
Yurdugül, Halil	Kategorik Değişkenler İçin İlişki ve Bağımlılık Ölçülerinin Karşılaştırılması, Yayınlanmamış yüksek lisans tezi, 1998
İnal, Ceyhan ve Günay, Süleyman 1993	Olasılık ve Matematiksel İstatistik, H.Ü. Fen Fakültesi Basımevi, Ankara
John, P. 1971	Statistical Design and Analysis of Experiments, Newyork.
Şimşek ve Yıldırım	Sosyal Bilimlerde Nitel Araştırma Yöntemleri, Seçkin Yayınları, 1999 Ankara

Halil Yurdugül
yurdugul@hacettepe.edu.tr

EK 1

$$E_{xx} = \sum\sum(x_{ij} - x_{i.})^2$$

$$T_{xx} = \sum\sum(x_{i.} - x_{..})^2$$

$$S_{xx} = \sum\sum(x_{ij} - x_{..})^2$$

$$E_{xy} = \sum\sum(x_{ij} - x_{i.})(y_{ij} - y_{i.})$$

$$T_{xy} = \sum\sum(x_{i.} - x_{..})(y_{i.} - y_{..})$$

$$S_{xy} = \sum\sum(x_{ij} - x_{..})(y_{ij} - y_{..})$$

$$E_{yy} = \sum\sum(y_{ij} - y_{i.})^2$$

$$T_{yy} = \sum\sum(y_{i.} - y_{..})^2$$

$$S_{yy} = \sum\sum(y_{ij} - y_{..})^2$$