



Makro İktisat Verilerinde Kayıp Verilerin Regresyona Dayalı En Yakın Komşu “Hot Deck” Yöntemi İle Tamamlanması

A.Mete ÇİLİNGİRTÜRK*

Dilek ALTAŞ**

Özet

Ülke verileri söz konusu olduğunda araştırmacılar çok boyutlu uzaklıkları kullanan kümeleme yaklaşımlarını tercih etmektedirler. Ancak karşılıklı bağımlılık içeren değişkenler arasında kayıp verilerin tamamlanmasında regresyon yönteminin kullanılması yaygındır. Bunun sebebi değişkenler arasındaki bağımlılık yapısının bozulmamasının amaçlanmasıdır. Bu durumda her iki analiz yaklaşımının da ortak ele alınarak hem göstergeler arasındaki bağımlılığın korunması hem de ülkeler arasındaki anlamlı uzaklıkların yok edilmemesini sağlayacak yöntemler tercih edilmelidir. Çalışmada regresyona dayalı en yakın komşuluk algoritmasının kullanılmasıyla elde edilen sonuçlar diğer yöntemlerle elde edilen sonuçlarla karşılaştırılmıştır.

Anahtar kelimeler: Kayıp veri tamamlama, İktisadi veri, Kümeleme

JEL Kodu:C82, C38

Abstract

The researchers prefer usually classification approaches which are depending on multidimensional distances to handle country data. However, it is used to impute the missing data with regression analysis at interdependent variables. The reason is to prevent the interdependency structure among variables. Therefore, there is a need for missing value imputation techniques, which will prevent the collinearity among economic variables and the significant distances among the countries. This research discussed the comparison between the results of hot-deck imputation with regression analysis and other imputation methods.

Key Words: Missing data imputation, Economical data, Classification

JEL Code:C82, C38

* Doç.Dr., Marmara Üniv., İ.İ.B.F., Ekonometri Bölümü, İstatistik Anabilim Dalı,
Bahçelievler-İstanbul, acilingi@marmara.edu.tr

** Doç.Dr., Marmara Üniv., İ.İ.B.F., Ekonometri Bölümü, İstatistik Anabilim Dalı,
Bahçelievler-İstanbul, dilekaltas@marmara.edu.tr

1. Giriş

Ekonometrik analizler ulusal ve uluslararası kurumlar ile hükümetlerin hazırlayıp düzenlediği göstergelere dayalı olarak yapılmaktadır. Ülkelere ait bazı bilgilerin elde edilememesi nedeniyle uluslararası panel veya yatay kesit verilerde kayıp veriler oluşmaktadır. Tek değişkenli tamamlama teknikleri ekonomik ve sosyal göstergeler arasındaki bağımlılık yapısından dolayı yetersiz kalmaktadır. Ülkelerin sınıflandırılmasına veya kümelenmesine yönelik araştırmalarda çok boyutlu uzaklıkları kullanan kümeleme yaklaşımları tercih edilmekte, ancak kayıp verilerin tamamlanmasında değişkenler arasında bağımlılık yapısı olduğundan regresyon yönteminin kullanılması tercih edilmektedir. Bunun sebebi, iktisadi ve sosyal göstergeler arasındaki bağımlılık yapısının bozulmamasının amaçlanmasıdır. Bu durumda her iki analiz yaklaşımının da ortak ele alınarak hem göstergeler arasındaki bağımlılığın korunması hem de ülkeler arasındaki anlamlı uzaklıkların yok edilmemesini sağlayacak yöntemler tercih edilmelidir.

Yapısal göstergeler, genel ekonomik alan, istihdam, yenilik ve araştırma, ekonomik reform, sosyal birleşme ve çevresel koşullar gibi altı alt başlığı kapsamaktadır. Avrupa konseyi, 2000 yılı Mart ayındaki Lizbon toplantısında, Avrupa Birliği gelecek 10 yıl için “daha iyi ve daha fazla iş, daha geniş sosyal birleşme ile sürdürülebilir ekonomik büyümenin dünyada en rekabetçi ve dinamik bilgiye dayalı ekonomi olması” şeklinde ifade edilen stratejik bir amaç belirlemiştir. Bu amaçla Eurostat’da söz konusu yapısal göstergelere ait verileri içeren siteler yer almaktadır. Aynı zamanda makro ekonomik göstergeleri düzenleyen diğer ulusal ve uluslar arası kuruluşlar da benzer çalışmalar yapmaktadırlar (Denk ve Hackl, 2003:306-310).

OECD, dünya çapında ekonomik çalışmalar yapmak amacıyla güvenilir veriler düzenlemektedir. Çalışmada OECD ülkelerinin ekonomik verilerinden yararlanılacaktır. Yapılan literatür araştırmasında ülke verilerinin düzenlenmesinde temel istatistiklerin kullanıldığı ortaya konulmuştur. Bu şekilde revize edilen verilerle önemli çalışmalar yapılmaktadır. Bunlardan ilk göze çarpanı küresel (genel) görünümün risklerini tahmin etmenin yanında, uluslararası makro ekonomik politikanın dağılımı ve etkilerini analiz etmek için OECD ekonomik bölümünün anahtar araçlarından biri, INTERLINK olarak adlandırılan makro ekonomik modeldir. Bölümün düzenli proje araştırmalarında model çeşitli şekillerde rol oynamaktadır. Bunlar,

- bireysel ülke projelerinin düzenlenmesine ve yorumlanmasına katkı sağlaması,
- dünya pazarında uyumlu ticaret projelerinin üretilmesi,
- alternatif ekonomik koşulların ve politik varsayımların sonuçlarının kısa ve orta vadede keşfedilmesi

şeklinde tanımlanmışlardır (Dalsgaard, André ve Richardson, 2001:32).

Söz konusu modellerde ve yapılan analizlerde eksik verilerin yer alması elde edilen sonuçların doğruluğunu etkilemektedir. Doğruluk, temelde eksik veri sayısı ve serilerdeki tartışılabilir değerlerin sayısına bağlıdır. Düzenlenen yeni göstergeler, ayrıca aşırı değerlerin sayısına, revizyon analizlerine ve düzensiz dalgalanmalara da önem vermektedir. Eksik veri, aylık, üçer aylık ya da yıllık herhangi bir t zamandaki değerinin bilinmemesi olarak tanımlanmaktadır. Ülke verilerini düzenleyen kuruluşların yayınlarında ekonominin ödemeler dengesi, üretici ve tüketici anketleri, tüketici fiyatları, dış ticaret, endüstri, ticaret ve hizmetler, işgücü piyasası, para ve finansal göstergeler gibi alt gruplarına ait bazı verilerin eksik olduğu belirlenmiştir. 2008 yılı Şubat ayına ait verilerde toplam serinin 594 (%1) tanesinin eksik olduğu görülmektedir. Bunların 240 tanesinin 10 veriden daha fazla, 190 tanesinin 2 ile 10 arasında, 164 tanesinin sadece bir adet eksik veri içerdiği tespit edilmiştir (Eurostatistics-2008:1-23).

2. Metodoloji

Eksik verilerin tamamlanmasında temel istatistikleri kullanan basit yöntemlerden çeşitli analizleri temel alan algoritmalara kadar farklı uygulama ve yöntemler geliştirilmiştir. Bu ihtiyacın temel sebebi incelenen verilerin dağılım özelliklerinin bozulmamasını sağlamaktır.

Eksik verilerin ele alınması ile ilgili temel yaklaşımlar hiçbir verinin eksik olmadığı gözlemlerin kullanılması (Afifi ve Elashoff, 1966: 595-604; Little ve Rubin,1986) ve korelasyona dayalı yöntemlerde eksik verisi bulunmayan değişken çiftlerinin kullanılması (Little ve Rubin, a.g.e.) şeklinde ortaya çıkmışlardır. Model tabanlı olmayan eksik veri tamamlama yaklaşımları olarak ortalama ve kesikli verilerde medyan değerinin kullanılması yaygındır. Bu yöntemler dağılımı homojenliği artırıcı şekilde bozmakta ve gözlemlerin küme özelliklerini dikkate almamaktadırlar (Hawkins ve Merriam, 1991:21-24). Verilerin belirli kümeler veya sıralamalara göre girilmesi durumunda “en yakın komşu” gözlemlerin ortalama veya medyan değerinin kullanılması tercih

edilmektedir (Hastie, v.d., 1999). Yakın zamanda gelişen bu yaklaşım “hot deck imputation” olarak tanımlanmakta; diğer veri tabanlarının veya merkezi eğilim ölçülerinin kullanılmasında ise “cold deck imputation” adını almaktadır (Fogarty,2001:8). Model tabanlı eksik veri tamamlama yöntemlerinde ise zaman serilerinde eksik verinin komşularına göre yarı ortalamalar yönteminin kullanılması veya trend denklemi yardımı ile interpolasyon yapılması mümkün olmaktadır. Genel olarak literatür incelendiğinde eksik veri tamamlama yöntemleri Tablo-1’de verildiği şekilde özetlenebilir (Wasito, 2003:9-28).

Tablo 1 Eksik Veri Tamamlama Yöntemleri

Tahmin Kuralları	Basit	Merkezi Eğilim İle Tamamlama	Gerçek varyans $(n_k-1)/(n-1)$ oranında daha küçük tahmin edilerek sapmalı tahmincilerle sebep olur.
		Hot / Cold Deck Tamamlama	Yakın komşuların değerinin ikamesidir. Kavramsal ve teknik olarak basit, ölçek tipini bozmayan yaklaşımdır.
Çok Değişkenli	Regresyon ile Tamamlama	Eksik olmayan verilerden yararlanarak eksik veri içeren değişken için koşullu beklenen değer tahmini yapılmasını sağlar. Varyans küçük tahmin edilir. Eksik verilerin oranı ve dağılımından etkilenir.	
	Karar Ağacı Analizi Yaklaşımı	Eksik olmayan verilerden yola çıkılarak CHAID (kategorik) veya CART (nicel) yöntemleri ile koşullu dağılım belirlenerek uygun ortalama tamamlama yöntemi kullanılır.	
	Yapay Sinir Ağları İle Tamamlama	Parametrik ve doğrusal olmayan regresyon işlevi gösteren yaklaşımda öğrenme modelinin uyum iyiliği her zaman değerlendirilememektedir.	
Maksimum Benzerlik Tahminleri	Kesin istatistiksel bir modelin verinin tahmini ve eksik öngörülmesini amaçlar.	EM- Algoritmaları İle Tamamlama	Ad-hoc değerler ile dağılımın parametreleri tahmin edilerek, bu dağılımdan istenen konverjans seviyesine ulaşılan kadar tekrar eksik veriler için tamamlama yapılır.
		Markov Zinciri MC İle Çoklu Tamamlama	Tahmindeki belirsizlik ölçülebilmektedir. Ancak, yöntem pahalı ve karmaşık olup, çoklu normal dağılım varsayımı gerekmektedir. Ayrıca, hesaplamaların konverjansı açık değildir.
		Tam Bilgi Maksimum Benzerlikler	Eksiksiz veri setinden ortalama vektörü ve kovaryans matrisi hesaplanarak çoklu normal dağılıma sahip yığın için objelere ait log-benzerlik fonksiyonu elde edilir. Güçlü bir yöntem olup büyük yığınların kullanımına ihtiyaç duyar.

En Küçük Kareler Yaklaşımı	Parametrik olmayan bir yaklaşım olup, ana faktörün belirlenmesine dayanır.	Eksik Verisiz Model Yaklaşımı	Temel bileşenler analizi uygulamalarında eksik veri problemini çözmek amacı ile geliştirilmiştir. Tek ve daha çok boyutlu uzaylarda verinin basitleştirilerek tamamlama yapılmasına yardımcı olur. Çoğu durumda anlamlı hata görülmektedir.
		Tamamlanmış Veri Modeli Yaklaşımı	Önceki yaklaşımdan farklı olarak, eksik veriler ad-hoc değerler ile tamamlandıktan sonra ana bileşen elde edilecek şekilde ad-hoc değerlerin değiştirilmesi sağlanır. Eksiksiz veri modeli yaklaşımının sonuca ulaşmadığı durumda daha başarılı olabilmektedir. Ancak bu yöntem daha yavaş çalışmaktadır.

Gediga ve Düntsch (2003: 1-13) verinin kendisinden yola çıkılarak eksik verilerin tamamlanmasını sağlayan bir algoritma geliştirmişlerdir. Bu algoritma sayısal olmayan kurallara bağlanmıştır. Bu çalışmalarında belirli değerleri alan değişkenlerde eksik olmayanların arasındaki benzerliklerden yola çıkılarak eksik verileri benzer özelliklere sahip birimlerin değerleri ile tamamlamaktadır. Yakın komşuların belirlenmesinde regresyon analizi ile koşullu beklenen değerlerin kullanılması da mümkün olmaktadır (Laaksonen, 2000: 65-71). Bu yöntemin anketlere dayalı ölçekler için oldukça performans sağladığı belirtilmiştir. Benzer şekilde yakın komşulara göre eksik verilerin tamamlanmasında değişken seti ilgili değişken gruplarına ayrılarak eksik verilerin tamamlanması aşamalı olarak da gerçekleştirilmektedir (Northolt, 2000: 239-242).

3. Uygulama

Çalışmanın amacı, OECD ülkelerine ait 2005 yılı ekonomik verilerinden (World Development Indicators 2006) yola çıkarak gözlenen eksik verilerin regresyona dayalı en yakın komşu yöntemi ile kayıp verilerin tamamlanmasının ortalama ile tamamlama ve regresyon ile tamamlama yöntemlerine göre daha uygun olacağını göstermektir. Regresyon ile eksik veri tamamlamada stokastik yaklaşım kullanılmış ve tahmin edilen değere hata terimleri dağılımına göre tesadüfi hata eklenmiştir.

Regresyona dayalı eksik veri tamamlama aynı zamanda eksik verilere koşullu beklenen değer atama olarak tanımlanmaktadır. Mevcut eksik olmayan verilere göre modeller kurularak, eksik verilerin tamamlanması öngörülmektedir. Ancak iktisadi veriler söz konusu olduğunda makro iktisat teorisine uygun modellerin kurulması gerekliliği ortaya çıkmaktadır. Bu nedenle değişkenlerin tahmin değerleri doğrusal regresyon modelleri kullanıldığında

İktisadi gözlemler ile tutarlı çıkmamaktadır. Diğer taraftan bağımlı bir değişkenin tahmin edilmesinde kullanılan açıklayıcı değişkende de eksik verileri olması durumunda yöntem tek başına sonuç vermemektedir.

OECD verilerinden 177 ülkeye ait birbirleri ile modellenebilecek 5 iktisadi gösterge seçilmiştir. Bu değişkenlerden GSYİH ($X_{i,1}$) ve Net Dış Mamul Ticareti ($X_{i,2}$) göstergeleri eksik veri içermemektedir. Mevduat Faiz Oranı (%) ($X_{i,3}$) 144, Nihai Tüketim Harcamaları ($X_{i,4}$) 146 ve Toplam Rezervler ($X_{i,5}$) 164 gözlem içermektedir. Eksik veri tamamlamada kullanılan algoritma şu şekildedir:

1. Eksik verili değişkenler için tam veriye sahip değişkenler ile en uygun doğrusal regresyon modelleri kurulur,
2. Açıklanma oranı (R^2) en yüksek modelin bağımlı değişkeni için tahmin değerleri hesaplanır,
3. Bağımlı değişkenin tahmin değerleri küçükten büyüğe doğru sıralanır,
4. Bağımlı değişkenin eksik verisi tahmini değerine en yakın olan gözlemin orijinal değeri ile tamamlanır,
5. Eksik veri içeren değişken kalmayana kadar birinci aşamaya dönülür.

Analizlerde ikinci aşama için elde edilen regresyon modellerine ait sonuçlar aşağıdaki tabloda verilmektedir.

Tablo 2 Kullanılan Regresyon Modelleri

Döngü	Bağımlı Değişken	Açıklayıcı Değişkenler	R^2	F-anlamlılık
1	$X_{i,3}$	$X_{i,1}, X_{i,2}$	0,014	0,364
1 (kullanıldı)	$X_{i,4}$	$X_{i,1}, X_{i,2}$	0,989	0,000
1	$X_{i,5}$	$X_{i,1}, X_{i,2}$	0,584	0,000
2	$X_{i,3}$	$X_{i,1}, X_{i,2}, X_{i,4}$	0,017	0,493
2 (kullanıldı)	$X_{i,5}$	$X_{i,1}, X_{i,2}$	0,584	0,000
3 (kullanıldı)	$X_{i,3}$	$X_{i,1}, X_{i,2}, X_{i,4}, X_{i,5}$	0,019	0,622

Çalışmada regresyona dayalı en yakın komşuluk algoritmasının kullanılmasıyla elde edilen sonuçlar diğer yöntemlerle elde edilen sonuçlarla karşılaştırılacaktır.

Bu amaçla öncelikle kayıp veri içeren veri seti olmak üzere, ortalama Regresyon ve Hot Deck tamamlama yöntemleri ile tamamlanan veri setlerine

ilişkin betimsel istatistikler aşağıdaki tablodaki gibidir. GSYİH (Xi,1) ve Net Dış Mamul Ticareti (Xi,2) göstergeleri eksik veri içermediğinden değerlendirilmemiştir.

Tablo 3 Farklı Yöntemlere İlişkin Betimsel İstatistikler

	Betimsel İstatistikler	Mevduat Faiz Oran (%)	Nihai Tüketim Harcamaları	Toplam Rezervler
Tamamlanmamış Veri Seti	<i>Gözlem Sayısı</i>	144	146	164
	<i>Kayıp Veri Sayısı</i>	33	31	13
	<i>Ortalama</i>	6,2504	114699207575,3425	23962480408,5366
	<i>Standart sapma</i>	9,27114	314375733302,19240	86101100423,86730
	<i>Cronbach Alpha</i>	0,702		
Ortalama. Yöntemi	<i>Gözlem Sayısı</i>	177	177	177
	<i>Kayıp Veri Sayısı</i>	0	0	0
	<i>Ortalama</i>	6,2503	114699207575,3420	23962480408,5368
	<i>Standart sapma</i>	8,35689	285349215326,06350	82860236843,39110
	<i>Cronbach Alpha</i>	0,216		
Regresyon Yöntemi	<i>Gözlem Sayısı</i>	177	177	177
	<i>Kayıp Veri Sayısı</i>	0	0	0
	<i>Ortalama</i>	6,1252	127738679495,7209	22482945996,9343
	<i>Standart sapma</i>	8,58550	359409521635,52390	83091817780,72160
	<i>Cronbach Alpha</i>	0,421		
Hot Deck Yöntemi	<i>Gözlem Sayısı</i>	177	177	177
	<i>Kayıp Veri Sayısı</i>	0	0	0
	<i>Ortalama</i>	5,9405	127908581853,1074	22371207553,6723
	<i>Standart sapma</i>	8,53113	364130218919,18970	83059753714,36280
	<i>Cronbach Alpha</i>	0,437		

Tablodaki sonuçlar incelendiğinde, ortalama değerlerinin tamamlanmamış veri setine göre daha düşük çıktığı, en düşük ortalamının üç iktisadi değişken için de 3. yöntemle ait olduğu, ortalama yöntemi ile tamamlanan veri setinde standart sapma değerlerinin en küçük çıktığı tespit edilmiştir. Eksik verilerin ortalama yöntemi ile tamamlanması durumunda değişkenliğin daha az çıkması doğal bir sonuçtur. Çünkü eksik veriler ortalama yöntemi ile tamamlandığında sapmalar küçük çıkacak dolayısıyla bu da standart sapmanın düşük çıkmasına yol açacaktır. İktisadi değişkenler arasında anlamlı bir ilişkinin olup olmadığının belirlenmesi amacıyla hesaplanan Cronbach Alpha katsayısı tamamlanmamış veri setinin dışında en düşük ortalama tekniğinde, en yüksek ise Hot Deck Yöntemi ile elde edilmiştir. Bu da üçüncü yöntemin diğer yöntemlere göre veri setinin iç tutarlılığını daha iyi koruduğunu, diğer bir ifade ile değişkenler arası ilişki yapısını diğerlerine göre daha az bozduğunu göstermektedir.

Çalışmada yöntemlerin karşılaştırılması amacıyla Kümeleme Analizi uygulanarak, her veri seti için elde edilen küme yapıları anlamlılıkları açısından incelenmiştir. Uygulama 177 adet veri içerdiğinden uygun küme sayısının belirlenmesinde $k = \sqrt{n/2}$ formülünden (Tatlidil, 2002:341) yararlanılmış ve $k = \sqrt{177/2} \approx 9$ olarak elde edilmiştir. Üç yöntem ile tamamlanan veri setleri için kümeleme analizi sonuçları aşağıdaki tabloda verilmiştir.

Tablo 4 Farklı Yöntemlere İlişkin Kümeleme Analizi Sonuçları

	Tamamlanmamış Veri Seti	Ortalama Yöntemi	Regresyon Yöntemi	Hot Deck Yöntemi
1. KÜME	Avusturya, Arjantin, Belçika, Danimarka, Finlandiya, Yunanistan, Hong Kong, Endonezya İrlanda, Norveç,Portekiz Polonya, Suudi Arabistan, Güney Afrika,İsveç, Türkiye	Avusturya, Arjantin, Belçika, Danimarka, Finlandiya, Yunanistan, Hong Kong, Endonezya, İran, İrlanda, Norveç, Portekiz, Polonya, Suudi Arabistan, Güney Afrika, İsveç, İsviçre, Tayland, Türkiye	Solomon Adaları	Avusturya, Suudi Arabistan, Norveç, İrlanda, İsveç, Belçika, İsviçre, Endonezya, Hong-Kong, Arjantin, Finlandiya, Tayland, İran, Polonya, Güney Afrika, Portekiz, Yunanistan, Türkiye
2. KÜME	Fransa	Fransa, İngiltere	Fransa, İngiltere	Fransa, İngiltere
3. KÜME	Çin	Çin	Çin	Çin
4. KÜME	İspanya	Japonya	Japonya	Japonya
5. KÜME	Almanya	Almanya	Almanya	Almanya
6. KÜME	Brezilya, Hindistan, Kore, Meksika, Hollanda,Rusya	Brezilya, Kanada, Hindistan, Kore, Meksika, Hollanda,Rusya, İspanya,	Brezilya, Avustralya, Kanada, Hindistan, Kore, Meksika, Hollanda, Rusya, İspanya	İspanya,Avustralya, Hindistan,Brezilya, Hollanda, Kore, Kanada, Rusya
7. KÜME	İtalya	İtalya	İtalya	İtalya
8. KÜME	İngiltere	Amerika	Amerika	Amerika
9. KÜME	Diğer Ülkeler	Diğer Ülkeler	Diğer Ülkeler	Diğer Ülkeler

Tabloda incelendiği gibi her yöntem için küme yapılarında çok önemli farklılık olmadığı görülmektedir. Eksik veri içeren veri seti dışında üç yöntemde

de Amerika, Çin, Japonya, Almanya ve İtalya'nın tek başına küme oluşturduğu, gelişmiş Avrupa ülkeleri olan Fransa ve İngiltere'nin aynı kümede yer aldığı, Solomon Adalarının diğer yöntemlerde değil, sadece regresyon yönteminde ayrı kümede yer aldığı, Türkiye'nin komşu ülke olan Yunanistan ile İran ve Avrupa Birliği ülkeleri Belçika, Polonya, Portekiz gibi ülkelerle aynı kümede yer aldığı tespit edilmiştir.

Sonuç ve Öneriler

Çalışmada kayıp veri probleminin çözümü için geliştirilen teknikler özetlenmeye çalışılmıştır. Bu tekniklerin güçlü ve zayıf yönleri tablolaştırılarak açıklanmıştır. Uygulamada OECD ülkeleri 2005 yılı ekonomik verilerinden yararlanılarak gözlenen kayıp verilerin regresyona dayalı en yakın komşu yöntemi (Hot Deck) ile tamamlanmasının ortalama ve regresyon yöntemlerine göre daha uygun olduğunu göstermek amaçlanmıştır. Bu amaçla 177 ülkeye ait birbirleri ile modellenebilecek ve kayıp veri içeren GSYİH, Net Dış Mamul Ticareti, Mevduat Faiz Oranı, Nihai Tüketim Harcamaları ve Toplam Reservler olmak üzere beş iktisadi gösterge seçilerek kayıp veri tamamlamada kullanılan algoritmalar uygulanmıştır.

Kayıp veri içeren veri seti, Ortalama, Regresyon ve Hot Deck Yöntemleri ile tamamlanan veri setleri öncelikle betimsel olarak incelenmiştir. Bunun sonucunda Hot Deck yönteminin diğer yöntemlere göre veri setinin iç tutarlılığını daha iyi koruduğu, diğer bir ifade ile değişkenler arası ilişki yapısını diğerlerine göre daha az bozduğu tespit edilmiştir. Ayrıca her bir veri setine Kümeleme Analizi uygulanarak küme yapılarının anlamlılıkları da incelenmiştir.

Çalışma, AB ülkelerine ait makro iktisat verileri ve panel verileri kullanarak ekonometrik ve çok değişkenli istatistiksel analizler yapacak teorisyen ve pratisyenlerin kayıp veri sorunlarına ışık tutabilir.

Kaynakça

Afifi, A.A., R.M., Elashoff, (1966), "Missing Observations in Multivariate Statistics I: Review of the Literature", Journal of American Statist. Assoc., 61 , 595-604.

Dalsgaard, T., C., Andre ve P., Richardson (2001), "Standard Shocks In The Oecd Interlink Model" Economics Department Working Papers No. 306, 32

Denk, M., P., Hackl (2003), "Data Integration and Record Matching: An Austrian Contribution to Research in Official Statistics", Austrian Journal of Statistics, Volume 32 , Number 4, 305-321

Eurostatistics (2008), Data for Short-term Economic Analysis, Volume 2, 1-23.

Fogarty, D.J. (2001), "Multiple Imputation as a Missing Data Approach to Reject Inference on Consumer Credit Scoring", Affiliation, , s:8.

Gediga,G.; I., Duntsch (2003), "Maximum Consistency of Incomplete Data via Non-invasive Imputation", Artificial Intelligence Review, Volume 19 , Issue 1, 93-107

Hastie, T.; R. Tibshirani, G. Sherlock, M. Eisen, P. Brown, ve D. Botsein (1999), Imputing missing values for gene expression arrays. Technical report, Divison of Biostatistics, Stanford University.

Hawkins, M.R., V.H., Merriam, (1991), "An Overmodeled World", Direct Marketing, 21-24.

Laaksonen, S. (2000), "Regression-based nearest neighbour hot decking", Comput. Statist. 15, 65–71.

Little, R., D., Rubin (1986), "Statistical Analysis with Missing Data", Wiley, New York .

Northolt, E.S.(2000), "Hot deck imputation in the Dutch Structure of Earnings Survey", Proceedings ETK European Statistical Laboratory, 239-242.

Tatlıdil, H., "Uygulamalı Çok Değişkenli İstatistiksel Analiz", Ankara, 2002

Wasito, I. (2003), "Least Squares Algorithms with Nearest Neighbour Techniques for Imputing Missing Data Values", Doktora Tezi, University of London, 9-28.