

EKSİK GÖZLEM DEĞERLERİNE SAHİP OECD ÜLKELERİNİN BEBEK SAĞLIĞI İLE İLGİLİ ANALİZİNDE YERİNE KOYMA YÖNTEMLERİNİN KULLANILMASI

USING IMPUTATION METHODS IN THE ANALYSIS OF INFANT HEALTH IMPORTANCE FOR OECD COUNTRIES WHICH HAVE MISSING OBSERVATIONS

Gamze ÖZEL, Nihal ATA

Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü

ÖZET: Birçok veri kümesinde bazı değişkenlere ait gözlem değerleri kayıt edilememektedir. Bu durumda genellikle eksik gözlemlerin bulunduğu değişkenler ihmal edilerek analiz yapılmaktadır. Ancak bu gözlemlerin çalışmadan çıkartılmaları bilgi kaybına neden olur ve çalışma yetersiz kalır. Varyans analizinde, çok etkenli deney düzenlerinde, örnekleme çalışmalarında ve çok değişkenli istatistiksel analiz gibi çeşitli alanlarda eksik gözlemler ile ilgili birçok çalışma yapılmıştır. Çok değişkenli istatistiksel analizde kullanılan ortalama ve regresyon yöntemini içeren yerine koyma (imputation) yöntemi bunlardan birisidir. Bu çalışmada, OECD'ye üye ülkelere ait eksik gözlemlerin tahmini için ortalama ve regresyon yöntemleri kullanılarak eksik gözlemlerin tahmin edilmesi, çok değişkenli istatistiksel yöntemler ile ülkelerin bebek sağlığına verdiği önem bakımından sınıflandırılması ve çok boyutlu uzaydaki yerlerinin belirlenmesi amaçlanmıştır. Ayrıca Türkiye'nin OECD ülkeleri içindeki sıralamasına bağlı olarak elde edilen sonuçlar Türkiye için yorumlanmış ve Türkiye'nin OECD ülkeleri içindeki yeri belirlenmiştir.

Anahtar kelimeler: Eksik gözlem, Ortalama yöntemi, Regresyon yöntemi, Çok değişkenli istatistiksel yöntemler, Bebek sağlığı.

ABSTRACT: In many data sets some value of variables cannot be recorded. In these cases, the analysis is carried on by ignoring the variables which have missing observations but if these observations are important for the research; their extraction causes the loss of information. There are many studies about missing data in factorial designs, ANOVA, sampling survey studies and multivariate analysis, etc. In this paper we used mean imputation and regression imputation methods to estimate missing observations of countries which are the members of OECD. Then by using multivariate statistical methods, classification of countries according to their attention level of infant health and determination of their location in multidimensional space are aimed. Also, depending on the arrangement of Turkey in OECD countries, results are interpreted for Turkey and determined its place in OECD countries.

Keywords: Missing observation, Mean imputation, Regression imputation, Multivariate statistical method, Infant health.

1. Giriş

Dünya Sağlık Örgütü'nün yaptığı araştırmalara göre dünyada her yıl doğan 8 milyon bebek 1 yaşına gelmeden ölmektedir. Ülkemizde ise, her yıl 2 milyona yakın kadın gebe kalmakta, yılda 1.481.000 canlı doğumdan 48.280'inin 1 yaşına bile gelmeden öldüğü bilinmektedir. Gelişmekte olan ülkelerde her üç bebekten biri sağlıklı barmaktan yoksundur; her beş bebekten birine temiz içme suyu bulunamamaktadır; her yedi bebekten biri en temel sağlık hizmetlerine ulaşamamaktadır. Yoksul

yörelere sağlık hizmetleri veren kuruluşlar yetersizdir; daha da kötüsü insanların bu kuruluşlara ulaşabilme olanakları yoktur. OECD ülkeleri içinde ise, ekonomik ve sosyal koşullardaki genel gelişmelerin yanı sıra, çocuk aşılardan yararlanma dâhil, doğum sonrası sağlık hizmetlerindeki düzelmeler sayesinde, son on yıl içerisinde bebek ölüm oranlarında dikkate değer bir düşüş kaydedilmiştir. Portekiz’de 1970 yılından bu yana bebek ölümleri %90 azalmıştır. Türkiye, Meksika, İtalya, İspanya ve Yunanistan’da da bebek ölüm oranlarında büyük düşüşler sağlanmıştır. 2002 yılında OECD ülkeleri arasındaki en düşük bebek ölüm oranlarına sahip ülkeler İzlanda, Japonya ve bazı İskandinav ülkeleridir.

Bir çalışmada eksik gözlemlerin bulunması durumunda, olması gereken yapı bozulacak ve tüm gözlemler üzerinden parametre tahminleri yapılamayacaktır. Bu durumda öncelikle eksik gözlem sorununun ortadan kaldırılması gerekmektedir. Değişken sayısının çok olması durumunda eksik gözlem sorununu gidermek için geliştirilen yöntemlerden bazıları tam kayıtlı ünite tabanlı yöntem, yerine koyma yöntemi, ağırlıklandırma yöntemi ve en çok olabilirlik yöntemidir. Bu yöntemler üzerine ilk çalışmalar Wilks (1932), Anderson (1957), Afifi ve Elashoff (1966), Hocking ve Smith (1968), Hartley ve Hocking (1971) tarafından yapılmıştır. Eksik gözlem üzerine yapılan çalışmalar Little ve Rubin (1987), Schafer (1997), Graham ve Hofer (2000), Schafer ve Graham (2002) ve Graham et al. (2003) tarafından sürdürülmektedir.

Tam kayıtlı ünite tabanlı yöntemde, eksik gözleme sahip üniteler çıkartıldıktan sonra analiz sadece tam gözlemler üniteler üzerinden yapılmaktadır. İstatistiksel paket programlarda genellikle bu yöntem kullanılmaktadır. Ancak bu yöntem az sayıda eksik gözlem içeren veri kümeleri için uygundur. Ayrıca bu yöntemin kullanılması yanlış tahmin değerlerine ve bilgi kaybına neden olmaktadır. Tam kayıtlı ünite tabanlı yöntemi genellikle veri kümesindeki eksik gözlem oranının 0,05’ten küçük olması durumunda tercih edilmektedir. Ağırlıklandırma yönteminde, eksik gözlem içermeyen önceki örnekleme çalışmalarından bir ağırlık değeri elde edilmekte ve bu ağırlık değeri eksik gözlemlerin tahmininde kullanılmaktadır. En çok olabilirlik (ML, Maximum Likelihood) yönteminde, gözlemlerin çok değişkenli normal dağılımdan geldiği varsayılarak eksik gözlemlerin tahmini yapılmaktadır. Ortalama yöntemi ve regresyon yöntemini içeren yerine koyma yöntemi ise, eksik gözlemlerin tam gözlemlerle değişkenler yardımıyla tahminine dayanmaktadır (Little ve Rubin, 1987).

Bu çalışmada yerine koyma yöntemleri ile 30 OECD ülkesine ait değişkenlerdeki eksik gözlemler tahmin edilmeye çalışılmıştır. Daha sonra çok değişkenli istatistiksel yöntemlerden temel bileşenler analizi kullanılarak ülkelerin bebek sağlığına verdiği önem bakımından sıralaması ve kümeleme analizi kullanılarak bu ülkelerin sınıflandırılması amaçlanmıştır. Ayrıca çok boyutlu ölçekleme yöntemi ile de ülkelerin çok boyutlu uzaydaki yerleri tespit edilmeye çalışılmıştır. Elde edilen sonuçlar Türkiye için yorumlanmıştır.

2. Genel Bilgiler

Çalışmada materyal olarak Türkiye’nin de aralarında bulunduğu 30 OECD (Organisation Economic Co-Operation and Development) ülkesi incelenmiştir. Her bir ülkeye ait değişkenler, UNICEF’in 2002 yılı verilerinden elde edilmiştir. Araştırma kapsamına alınan ülkeler Tablo 1’de verilmiştir:

Tablo 1. OECD'ye Üye Ülkeler

1.ABD	7.Finlandiya	13.İsviçre	19.Slovakya	25.Japonya
2.Almanya	8.Fransa	14.Kanada	20.Yunanistan	26.Meksika
3.Avusturalya	9.Hollanda	15.Kore	21.Çek Cum.	27.Norveç
4.Avusturya	10.İngiltere	16.Lüksemburg	22.İsveç	28.Portekiz
5.Belçika	11.İrlanda	17.Macaristan	23.İtalya	29.Türkiye
6.Danimarka	12.İspanya	18.Polonya	24.İzlanda	30.Y.Zellanda

Çalışmada kullanılan değişkenler uluslararası araştırmalarda bebek sağlığı ile ilgili olarak kullanılan göstergeleri içeren bir nitelik taşımaktadır. Yukarıda belirtilen ülkeleri bebek sağlığına verdiği önem bakımından inceleyebilmek için kullanılmasına karar verilen değişkenler aşağıdaki gibidir:

- X_1 : Her 1000 canlı doğumdan 1 yaşında ve altında ölen bebeklerin oranı
 X_2 : Çocuk felci aşısı yapılan 1 yaşındaki bebeklerin oranı
 X_3 : Sağlıklı suyun ulaştığı kitle oranı
 X_4 : Yeterli sağlık önlemlerinin ulaştığı kitle oranı
 X_5 : 2,5 kilodan daha az doğum ağırlığına sahip bebeklerin oranı
 X_6 : Toplam doğurganlık hızı (Bir kadının doğurganlık çağının sonuna geldiğinde sahip olacağı ortalama çocuk sayısıdır.)
 X_7 : 15-49 yaş arasındaki kadınlardan gebeliği önleyici hap yada araç kullananların oranı

Değişken sayısının çok olması durumunda, gösterim kolaylığı sağlamak için çok değişkenli istatistiksel analizde vektör ve matrislerden yararlanılmaktadır. Veri matrisinin her bir satırı ünite olarak isimlendirildiğinde herhangi ünitenin p tane özelliğine ilişkin değerleri,

$$X'_i = [x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip}] \quad i = 1, \dots, n \text{ için} \quad (1)$$

biçimindeki sütun vektörü ile gösterilmektedir. Buna göre, n tane ünitenin p tane özelliğini gösteren veri matrisi aşağıdaki gibidir:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix} \quad i = 1, \dots, n \text{ ve } j = 1, \dots, p \text{ için} \quad (2)$$

Eşitlik (2)'de $X_{n \times p}$ biçiminde tanımlanan matriste ilk indis gözlem numarasını, ikinci indis ise değişken numarasını göstermektedir.

2.1. Eksik Gözlem Tahmini İçin Kullanılan Yerine Koyma Yöntemleri

2.1.1. Ortalama Yöntemi

Ortalama yöntemi, pratik bir yöntemdir ve bu yöntem ile kabul edilebilir sonuçlar elde edilmektedir. Ancak değişkenler arasındaki ilişki kuvvetli ise elde edilen sonuçlar yanıltıcı olabilmektedir. Ortalama yönteminde her bir değişkenin mevcut gözlemleri üzerinden hesaplanan ortalama, ilgili değişkendeki eksik gözlemlerin yerine konulmaktadır (Rubin, 1976).

Değişken sayısının birden fazla olması durumunda tüm değişkenlerin ortalamaları yardımıyla aşağıda verilen ortalamalar vektörü oluşturulur:

$$\bar{x}' = [\bar{x}_{.1} \quad \bar{x}_{.2} \quad \dots \quad \bar{x}_{.p}]. \quad (3)$$

Ortalama yönteminde n tane gözlemden m tane eksik x_{ij} gözlemi için eksik gözlem içermeyen üniteler üzerinden ortalama,

$$\bar{x}_{.j}^{(i)} = \frac{1}{n-m} \sum_{i=1}^{n-m} x_{ij} \quad j=1, \dots, p \quad \text{için} \quad (4)$$

biçiminde hesaplanır ve elde edilen bu değer ilgili değişkenin eksik gözlemlerinin yerine konulur. Eksik gözlem içeren tüm değişkenlerin ortalaması için,

$$\bar{x}'^{(j)} = [\bar{x}_{.1}^{(1)} \quad \bar{x}_{.2}^{(2)} \quad \dots \quad \bar{x}_{.p}^{(p)}] \quad (5)$$

biçiminde bir ortalamalar vektörü oluşturulur (Little ve Rubin, 1987).

2.1.2. Regresyon Yöntemi

Regresyon yöntemi ilk kez 1960 yılında Buck tarafından kullanılmıştır. Bu yüzden Buck's yöntemi olarak da bilinmektedir ve EM (Expectation-Maximization Algorithm) algoritmalarının özel bir biçimidir (Little, 1976).

Regresyon yönteminde her ünite için sırasıyla eksik gözlem içeren değişken bağımlı değişken olarak kabul edilerek regresyon katsayıları hesaplanır ve bu katsayılar yardımı ile eksik gözlemin bir tahmini elde edilir. Veri kümesinde aynı ünite de bir ya da birden fazla eksik gözlem bulunabilir. Bu durumda yöntem aşağıdaki gibi uygulanır:

(i) Tek Eksik Gözlemlili Ünite Durumu

X veri matrisinde p tane değişkene ait n gözlemin tam gözlemlili k ünitesi olsun. Veri matrisi eksik gözlemlili satırlardan oluşan kısım ve tam gözlemlili satırlardan oluşan kısım olmak üzere ikiye ayrılır. Daha sonra k tane gözlemden oluşan p değişkene ait veri grubu için ($X_{k \times p}$) diğer p-1 değişken üzerinden çoklu regresyon yapılır ve elde edilen regresyon katsayıları yardımı ile eksik gözlemler tahmin edilir. Bu amaçla kullanılan tahmin denklemi, j eksik gözlemi gösteren indis olmak üzere, aşağıdaki gibidir:

$$\hat{x}_j = \beta_0 + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1}. \quad (8)$$

(ii) Birden Fazla Eksik Gözlemlü Üniteler Durumu

Bir ünite için birden fazla değişkende aynı anda eksik gözlem olması durumunda X veri matrisinin m tam gözleme sahip ünitesi üzerinden eksik gözlemi olmayan değişkenler ile her eksik gözleme sahip değişken için çoklu regresyon analizi yapılır.

Regresyon yöntemi, iteratif bir yöntemdir. Öncelikle regresyon çözümlemesi ile veri kümesindeki tüm eksik gözlemler tahmin edilir. Daha sonra, tüm veri matrisi kullanılarak yeni tahmin denklemleri elde edilir. Bu tahmin denklemleri de yeni tahmin değerleri olan \hat{x}_{ij} hesaplamasında kullanılır. Buradan elde edilen yeni veri matrisi kullanılarak tekrar tahmin denklemleri ve \hat{x}_{ij} tahmin değerleri elde edilir. Bu süreç, adımlar sonrasında elde edilen değerler arasındaki fark en küçük düzeye gelinceye dek devam eder.

Regresyon yöntemi, ortalamalar yönteminden daha iyi sonuç vermektedir. Ancak diğer değişkenler tahmin edilecek değişken ile çok yüksek derecede ilişkili değil ise, regresyon yöntemi ortalama yöntemine eşit olmaktadır (Graham, et. al 2003).

2.2. Çok Değişkenli İstatistiksel Yöntemler

2.2.1. Temel Bileşenler Analizi

Çok değişkenli istatistiksel analizde, n tane nesneye ilişkin p tane değişken (özellik) incelenmektedir. Bu özelliklerden birçoğunun birbiri ile ilişkili ve p sayısının çok büyük olması sorun yaratmaktadır. Temel bileşenler analizi değişkenler arasındaki bağımlılık yapısının yok edilmesi ve/veya boyut indirgemek için uygulanmaktadır.

Temel bileşenler analizinde, araştırma kapsamına alınan değişkenlerin ölçü birimleri genellikle farklıdır. Tüm değişkenlerin ölçü birimleri aynı olsa da büyük varyansa sahip değişkenler temel bileşende daha büyük ağırlıklara, küçük varyansa sahip değişkenler ise daha küçük ağırlıklara sahip olur. Bu sakıncaları gidermek için değişkenlerin standartlaştırılması gerekmektedir. Bu nedenle çözümleme $Z_{p \times n}$ biçiminde tanımlanan standartlaştırılmış veri matrisi üzerinden gerçekleştirilmekte ve ilişkili z_{ij} değerlerinden, dönüştürme ile ilişkisiz y_{ij} değerlerine ulaşılmaktadır ($Y_{p \times n} = T'_{p \times p} Z_{p \times n}$).

Bartlett küresellik testi ile temel bileşenler analizi uygulanıp uygulanmaması gerektiğine karar verildikten sonra temel bileşen sayısı aşağıda verilen Kaiser Kuralını sağlayan en küçük m değerine bakılarak belirlenmektedir (Tatlıdil, 1996):

$$\sum_{j=1}^m \lambda_j / p \geq 2/3.$$

2.2.2. Kümeleme Yöntemi

Kümeleme yöntemi, uzaklık matrisinden birimleri ya da değişkenleri kendi içinde homojen ve birbirleri arasında heterojen gruplar oluşturmak için kullanılmaktadır. Uzaklık ölçüleri ya da benzerlik ölçüleri değişkenlerin ölçü birimlerine göre farklılıklar göstermektedir. Değişkenler oransal ya da aralıklı ölçekle ölçülmüş ise, öklid uzaklık ölçülerinden yararlanılmaktadır. Kümeleme yönteminde normal dağılım varsayımı prensipte kalmakta, uzaklık değerlerinin normalliği yeterli

görülmemektedir. Ayrıca bu çözümleme yönteminde kovaryans matrisine ilişkin herhangi bir varsayım bulunmamaktadır.

Kümeleme yöntemi, uzaklık matrisi ya da değişkenleri uygun gruplara ayırırken, grupları belirlemede izledikleri yaklaşımlara göre iki temel gruba ayrılmaktadır. Bunlar, aşamalı kümeleme yöntemi ve aşamalı olmayan kümeleme yöntemidir (Sharma, 1996).

Aşamalı kümeleme yönteminde, birimlerin benzerliklerini dikkate alarak belirli düzeylerde (küme uzaklık ölçüleri) birbirleri ile birleştirilmektedir. Kümeleme sürecinin başlangıcında her gözlem bir küme olarak alınır. Süreç sonunda ise, tüm gözlemler bir kümede toplanır. Aşamalı kümeleme yönteminde ağaç diyagramından (dendrogram) yararlanılarak grup sayısına karar verilir.

Aşamalı olmayan kümeleme yöntemi, küme sayısı konusunda bir ön bilgi varsa ya da araştırmacı anlamlı küme sayısına karar vermiş ise tercih edilmektedir. Ayrıca kuramsal dayanıklılıklarının daha güçlü olması da diğer bir tercih sebebidir (Sharma, 1996).

2.2.3. Çok Boyutlu Ölçekleme Yöntemi

Çok boyutlu ölçekleme yöntemi, n tane nesne (birey-gözlem) arasındaki uzaklık değerlerini kullanarak nesnelerin çok boyutlu uzaydaki konumlarına ilişkin görüntülerini ortaya koymak için kullanılmaktadır. Bu yöntemin amacı, nesnelerin yapısını mümkün olduğunca az boyutla orijinal şekle yakın bir biçimde elde etmektir.

Çok boyutlu ölçekleme yönteminde metrik ve metrik olmayan ölçekleme olmak üzere iki tür ölçekleme yönteminden yararlanılmaktadır. Metrik ölçekleme temel bileşenler analizine benzer bir yapıya sahiptir. Metrik olmayan ölçekleme yönteminde ise, uzaklıkların büyüklük bakımından sıralanmasıyla çözüme gidilmektedir. Ayrıca stress katsayısı hesaplanarak tolerans oranlarından elde edilen görüntünün gerçek görüntüye uygun olup olmadığına karar verilmektedir. Stress katsayısı, Kruskal tarafından geliştirilen tolerans oranlarından yararlanarak elde edilen görüntünün gerçek görüntü ile uygunluğunun bir ölçütüdür (Tatlıdil, 1996).

3. Türkiye'nin Bebek Sağlığı Bakımından OECD Ülkeleri Arasındaki Yerinin Belirlenmesi

Çalışmada kullanılan 30 ülkeye ait 7 değişkenin değerlerinden 30x7 boyutunda bir veri matrisi oluşturulmuştur. Veri matrisinde X1, X2, X6 değişkenleri tüm ülkeler için tam gözlem ve X3, X4, X5, X7 değişkenleri ise eksik gözlem içermektedir. Belçika, Japonya, Çek Cumhuriyeti, Danimarka, Finlandiya, Fransa, Almanya, Yunanistan, İzlanda, İrlanda, İtalya, Lüksemburg, Norveç, Polonya, Portekiz, İspanya, Kore ve Yeni Zelanda ülkelerinin X3, X4, X5 ve X7 değişkenlerine ait eksik gözlemleri ortalama yöntemi ile elde edilmiş ve sonuçlar aşağıdaki ortalamalar vektörü ile verilmiştir:

$$\bar{x}^{(j)} = [* * 0,976 0,947 0,06 * 0,7273]. \quad (10)$$

Burada (*), ilgili değişkenlerde eksik gözlem olmadığı için hesaplanmayan ortalamaları göstermektedir. Bu ortalama değerleri veri kümesinde ait oldukları

yerlerine konularak çok değişkenli istatistiksel yöntemlerin uygulanabilirliği sağlanmıştır.

Regresyon yönteminde ise, eksik gözlemleri tahmin edebilmek için öncelikle veri kümesi tam gözleme sahip ülkeler ve eksik gözleme sahip ülkeler ikiye ayrılmıştır. Bölüm 2.1.2 (i)'ye göre, eksik gözlem değeri içermeyen $k=12$ ülkeye ait alt veri matrisi üzerinden adimsal işlemlere başlanmıştır. Tek eksik gözlem içeren ülkelerin X4 veya X5 değişkenleri için ayrı ayrı 6 değişken üzerinden çoklu regresyon yapılmıştır. Bir ülke için birden fazla değişkende eksik gözlem olması durumunda, regresyon yönteminde eksik gözlemi olmayan değişkenler üzerinden her eksik gözleme sahip değişken için çoklu regresyon analizi yapılmıştır. Hesaplanan eksik gözlem değerleri son iki adımda birbirine oldukça yakın olduğundan iterasyonun 4. adımında sonlandırılmasına karar verilmiştir. Elde edilen parametre tahmin değerleri EK 1 ve bulunan eksik gözlem değerleri EK 2'de verilmiştir.

Her iki yöntem kullanılarak eksik gözlem tahmini yapılmış ve tam veri kümelerine ulaşılmıştır. OECD ülkelerini bebek sağlığına verdikleri önem bakımından sıralamak için temel bileşenler analizine geçmeden önce araştırma kapsamında incelenen 30 ülkeyi tanımlamaya yönelik değişkenlerin birimleri birbirinden farklı olduğundan standartlaştırma yapılmıştır. Standartlaştırılmış değişkenler kullanıldığından korelasyon matrisinden yararlanılmıştır.

Yapılan çalışmada ele alınan verilere temel bileşenler analizinin uygulanıp, uygulanmayacağını ve değişkenler arasındaki ilişkilerin önemli olup olmadığını görebilmek için Bartlett küresellik testi yapılmıştır. Bartlett küresellik testi sonuçları ortalama ve regresyon yöntemleri ile elde edilen veriler için sırasıyla ($\chi^2 = 116,302$; Sd=21; p-değeri=0,00) ve ($\chi^2 = 161,485$; Sd=21; p-değeri=0,00) olarak bulunmuştur. Buna göre, değişkenler arasındaki ilişkilerin önemli ve ilişki matrisi ile birim matris arasında fark olduğu 0,05 yanılma düzeyinde söylenebilir. Bu nedenle, değişkenler arasındaki bağımlılık yapısının yok edilmesi ve boyut indirgemek için temel bileşenler analizi uygulanması gerektiği sonucuna ulaşılmıştır.

Ortalama yöntemi ve regresyon yöntemi ile elde edilen veri kümeleri için temel bileşenler analizi yapıldıktan sonra Kaiser Kuralı dikkate alınarak oluşturulan temel bileşenler modelleri sırasıyla aşağıdaki gibidir:

$$\begin{aligned} y_1 &= -0,327z_1 + 0,012z_2 + 0,329 z_3 + 0,276z_4 - 0,134 z_5 - 0,271z_6 + 0,092z_7 \\ y_2 &= 0,202z_1 + 0,500z_2 - 0,029z_3 + 0,305 z_4 - 0,040z_5 + 0,045z_6 - 0,629z_7 \\ y_3 &= 0,017z_1 + 0,338z_2 + 0,012 z_3 - 0,260z_4 + 0,852z_5 + 0,048z_6 + 0,101z_7 \end{aligned}$$

$$\begin{aligned} y_1 &= -0,335z_1 + 0,021z_2 + 0,323 z_3 + 0,231z_4 - 0,081z_5 - 0,255z_6 + 0,159z_7 \\ y_2 &= 0,217z_1 + 0,514z_2 - 0,008z_3 + 0,290 z_4 + 0,049z_5 + 0,062z_6 - 0,583z_7 \\ y_3 &= 0,001z_1 + 0,239z_2 - 0,003 z_3 - 0,317z_4 + 0,871z_5 - 0,015z_6 + 0,028z_7. \end{aligned}$$

Bu modellerde ülkelere ait gözlem değerleri yerlerine konulduktan sonra OECD ülkelerinin bebek sağlığına verdikleri önem bakımından sıralamaları Tablo 2 ve Tablo 3'teki gibi elde edilmiştir:

Tablo 2. Ortalama Yöntemi ile Elde Edilen Veriler Üzerinden Sıralama

1. Avusturya	11. Hollanda	21. İzlanda
2. Macaristan	12. Danimarka	22. Lüksemburg
3. Polonya	13. İsveç	23. Norveç
4. Japonya	14. Finlandiya	24. Meksika
5. Slovakya	15. Fransa	25. Y.Zellanda
6. Portekiz	16. İtalya	26. Kanada
7. Belçika	17. Avusturalya	27. İrlanda
8. Çek Cum.	18. Yunanistan	28. Almanya
9. İngiltere	19. İsviçre	29. Kore
10. ABD	20. İspanya	30. Türkiye

Tablo 3. Regresyon Yöntemi ile Elde Edilen Veriler Üzerinden Sıralama

1. Macaristan	11. ABD	21. İsviçre
2. Avusturya	12. İtalya	22. Lüksemburg
3. Japonya	13. Danimarka	23. İzlanda
4. Polonya	14. Fransa	24. Norveç
5. Portekiz	15. İspanya	25. Y.Zellanda
6. Slovakya	16. Finlandiya	26. Kanada
7. Belçika	17. Yunanistan	27. Almanya
8. Çek Cum.	18. İsveç	28. İrlanda
9. İngiltere	19. Avusturalya	29. Kore
10. Hollanda	20. Meksika	30. Türkiye

Ortalama yöntemi ile elde edilen veriler ile yapılan sıralamada Avusturya, Macaristan ve Polonya'nın; regresyon yöntemi ile elde edilen veriler ile yapılan sıralamada ise Macaristan, Avusturya ve Japonya'nın bebek sağlığına en çok önem veren ülkeler olduğu görülmüştür. Her iki yöntem ile elde edilen verilere dayanarak Türkiye 30 OECD ülkesi arasında en son sıra yer almıştır.

Temel bileşenler analizi ile ülkelerin sıralamaları yapıldıktan sonra, EK 3'te verilen öklit uzaklık katsayıları kullanılarak aşamalı kümeleme yöntemi ile elde edilen anlamlı ve uygun sınıflandırmalar, Tablo 4 ve Tablo 5'teki gibidir. Ayrıca ülkeler arasında sınıflandırmalar yapabilmek için kullanılan ve öklit uzaklıklarına dayanan ağaç diyagramları da EK 4 ve EK 5'te verilmiştir.

Tablo 4. Ortalama Yöntemi ile Elde Edilen Veriler Üzerinden Sınıflama

Küme No	Ülkeler
1	İspanya, İtalya, Yunanistan, Belçika, Macaristan, Slovakya
2	Avusturalya, ABD, İngiltere, Fransa, Danimarka, Finlandiya, Hollanda, İsviçre
3	İrlanda, Norveç, Yeni Zellanda, Lüksemburg, İzlanda, İsveç
4	Almanya, Kanada
5	Portekiz, Japonya, Çek Cumhuriyeti, Polonya, Avusturya
6	Kore
7	Türkiye
8	Meksika

Tablo 5. Regresyon Yöntemi ile Elde Edilen Veriler Üzerinden Sınıflama

Küme No	Ülkeler
1	Fransa, Finlandiya, Danimarka, Norveç, İzlanda, Lüksemburg, İsveç
2	ABD, Avusturalya, Slovakya, Hollanda, İtalya, İspanya, İsviçre, Belçika, İngiltere, Macaristan
3	Yeni Zellanda, İrlanda, Almanya, Kanada, Yunanistan
4	Polonya, Avusturya, Portekiz, Japonya, Çek Cumhuriyeti
5	Kore
6	Türkiye
7	Meksika

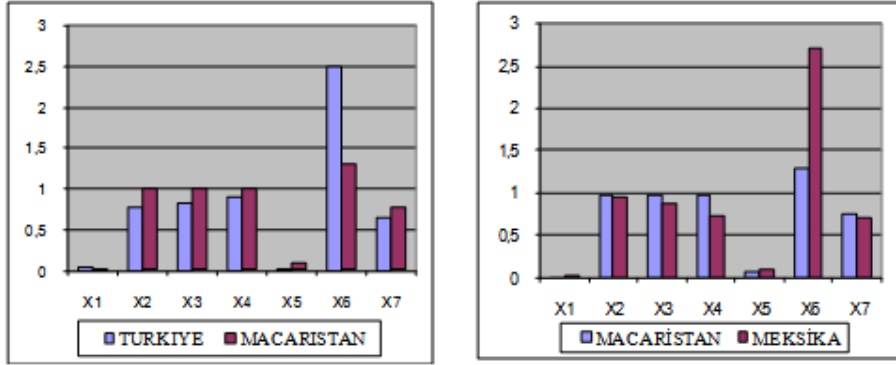
Her iki yöntem kullanılarak elde edilen veriler ile yapılan sınıflandırmalar sonucunda Kore, Türkiye ve Meksika tek başına ayrı kümeler oluşturduğu görülmüştür. Ancak diğer ülkelerin bulunduğu kümeler, yöntemler için farklılık gösterilmiştir.

Kümeleme yöntemi ile sınıflandırılan ülkelerin iki boyutlu uzaydaki yerlerinin tespiti için çok boyutlu ölçekleme yönteminden yararlanılmıştır. Tablo 6'da görüldüğü gibi, her iki yöntem için belirtme katsayısı oldukça büyüktür. Çalışmada stress katsayıları dikkate alınarak orta uyum olduğu söylenebilir. Bu durum, çalışmada elde edilen sonuçların kullanılabilir olduğunu göstermektedir.

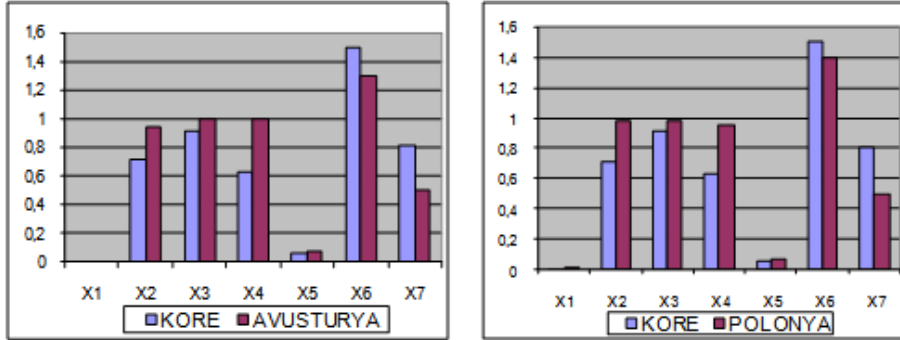
Tablo 6. Çok Boyutlu Ölçekleme ile Elde Edilen Katsayı Değerleri

	Ortalama yöntemi ile elde edilen veriler için	Regresyon yöntemi ile elde edilen veriler için
Belirtme Katsayısı	0,939	0,946
Stress Katsayısı	0,187	0,173

Ülkelerin iki boyutlu uzaydaki öklit uzaklıkları modeline göre elde edilen sonuçlar doğrultusunda daha anlamlı yorumlar yapabilmek için aşağıdaki grafikler çizilmiştir:

**Şekil 1. Ortalama Yöntemi ile Elde Edilen Verilerin 1. Boyuta Göre Birbirine en Uzak Ülkelerin Çizimleri**

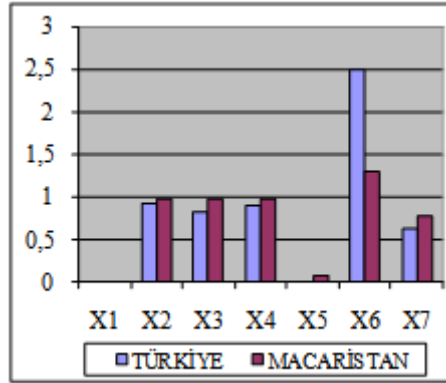
Şekil 1 incelendiğinde, Türkiye-Macaristan ve Macaristan-Meksika arasındaki farklılıkların büyük ölçüde ülkelerin toplam doğurganlık hızından kaynaklandığı görülmektedir.



Şekil 2. Ortalama Yöntemi ile Elde Edilen Verilerin 2. Boyuta Göre Birbirine en Uzak Ülkelerin Çizimleri

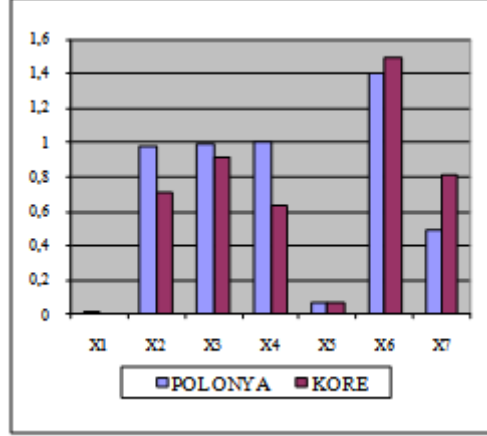
Şekil 2'ye göre, Kore ile Avusturya arasında farklılığa neden olan değişkenlerin sırasıyla yeterli sağlık önlemlerinin ulaştığı kitle oranı ve 15-49 yaş arasındaki kadınlarda gebeliği önleyici hap ya da araç kullananların oranıdır. Kore ile Polonya arasında farklılığa neden olan değişkenlerin ise, sırasıyla 15-49 yaş arasındaki kadınlarda gebeliği önleyici hap ya da araç kullananların oranı, yeterli sağlık önlemlerinin ulaştığı kitle oranı ve çocuk felci aşısı yapılan bir yaşındaki bebeklerin oranı olduğu görülmüştür.

Şekil 3 incelendiğinde, Türkiye ile Macaristan arasındaki farklılığın toplam doğurganlık hızı ve 15-49 yaş arasındaki kadınlarda gebeliği önleyici hap ya da araç kullananların oranı değişkenlerinden kaynaklandığı görülmüştür.



Şekil 3. Regresyon Yöntemi ile Elde Edilen Veriler İçin 1. Boyuta Göre Birbirine en Uzak Ülkelerin Çizimi

Aşağıda verilen Şekil 4'e göre, Polonya ile Kore arasında farklılığa neden olan değişkenlerin sırasıyla yeterli sağlık önlemlerinin ulaştığı kitle oranı, çocuk felci aşısı yapılan bir yaşındaki bebeklerin oranı ve toplam doğurganlık hızı olduğu görülmüştür.



Şekil 4. Regresyon Yöntemi ile Elde Edilen Veriler İçin 2. Boyuta Göre Birbirine en Uzak Ülkelerin Çizimi

4. Sonuç

Bu çalışmada, Türkiye'nin de aralarında bulunduğu OECD'ye üye 30 ülkeyi inceleyebilmek için UNICEF' in 2002 yılı verileri kullanılmıştır. Ancak veri kümesinde eksik gözleme sahip değişkenler olduğu için öncelikle bu eksik gözlemlerin tahminine çalışılmıştır. Bu amaçla yerine koyma yönteminin iki türü olan ortalama ve regresyon yöntemleri kullanılmıştır. Her iki yöntemden yararlanarak elde edilen veri kümeleri üzerinden çok değişkenli istatistiksel yöntemler ile sıralanan üye ülkeler, bebek sağlığına verdiği önem bakımından da sınıflandırılmıştır. Ayrıca üye ülkelerin iki boyutlu uzaydaki görünüşleri de elde edilmiştir. OECD ülkeleri içinde Türkiye bebek sağlığına verdiği önem bakımından son sırada yer almıştır. OECD ülkelerinde toplam sağlık harcamalarının GSMH içindeki oranı ortalama %7.2'dir. Aynı oranın Türkiye'de %3.7 olduğu ve yıllar boyunca önemli bir değişiklik göstermediği bilinmektedir. Sıralamada en iyi durumda olan Avusturya'da kişi başına yapılan sağlık harcaması miktarı 2225 dolar iken Türkiye'de 108 dolardır. Dolayısıyla çalışmada elde edilen sonuçların gerçek durumu yansıttığı ve Türkiye'nin sağlık sektörüne daha fazla önem vermesi gerektiği söylenebilir.

2003 yılı Türkiye Nüfus ve Sağlık Araştırması'na göre, ülkemizde her yıl canlı doğan 1000 bebekten 29'u bir yaşını tamamlamadan, 37'si beş yaşından önce ölmektedir. Bu rakamlar gelişmiş ülkelerle, hatta komşu ülkelerle karşılaştırılmayacak kadar yüksektir. Türkiye ekonomisi büyüklüğü bakımından dünyada ilk 20'ye girmesine karşılık bebeklerin sağlık durumu oldukça kötüdür. Örneğin, kişi başına geliri Türkiye'nin üçte birinden az olan Sri-Lanka'da, bebek ölüm oranı Türkiye'nin yarısı kadardır. Bu durum bebek ölümlerinin azaltılmasının yalnızca maddi olanaklar ve sağlık hizmetlerine değil, toplumun eğitim düzeyinin yükseltilmesine bağlı olduğunu göstermektedir. Bunun nedeni, bebek ölüm oranı aşırı doğum oranından genel yaşam standartlarının düşüklüğüne, sağlık imkanlarından kültürel özelliklere kadar uzanan geniş ve karmaşık bir faktörler dizisi ile ilişkili olmasındandır.

Referanslar

- AFIFI, A.A., ELASHOFF, R.M. (1966). Missing observations in multivariate statistics, I. Review of literature, *Journal of the American Statistical Association*, 61, 595-604.
- GRAHAM, J.W., CUMSILLE, P.E., ve ELEK-FISK, E. (2003). *Methods for handling missing data, Research Methods in Psychology, Vol. 2 of Handbook of Psychology*. New York, John Wiley & Sons, 87-114 ss.
- LITTLE, R.J.A. (1976) Inference about Means from Incomplete Multivariate Data, *Biometrika*, 63, 3, 593-604.
- LITTLE, R.J.A., RUBIN, D.B. (1987). *Statistical analysis with missing data*. New York, John Wiley and Sons.
- OECD in figures (2002). *Organisation for economic Co-operation and development*. Paris, France.
- OECD Göstergeleri (2005). “Sağlığa bakış” Erişim adresi: <<http://www.oecd.org/dataoecd/55/49/35636867.pdf>>.
- RUBIN, D.B. (1976) Inference and missing data. *Biometrika*, 63, 3, 581-592.
- SHARMA, S. (1996) *Applied multivariate techniques*. New York, Wiley.
- TATLIDİL, H. (1996). *Uygulamalı çok değişkenli istatistiksel analiz*. Ankara, Cem Web Ofset.
- YURDAKÖK, M. (2005), Dünyada ve ülkemizde çocuk sağlığı. *Çocuk Sağlığı ve Hastalıkları Dergisi*; 48, 203-205.

Ekler

EK 1. Regresyon Yöntemi ile Hesaplanan Parametre Tahminleri

Adımlar	Eksik gözlem içeren ünite sayısı	Her Adımdaki Bağımlı Değişkenler	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
1. adım	1	X4	-2,104	13,034	-0,066	3,239		-1,720	-0,052	0,071
		X5	-1,185	5,988	0,038	1,511	-0,340		-0,002	0,025
	2	X3	1,077	-4,828	-0,056			0,262	-0,018	0,014
		X4	1,385	-2,608	-0,247			-0,871	-0,110	0,117
	3	X3	1,087	-4,916	-0,058			0,268	-0,017	
		X4	1,469	-3,330	-0,263			-0,818	-0,099	
	X7	0,709	-6,162	-0,133			0,494	0,099		
2. adım	1	X4	-3,498	19,269	0,025	4,518		-2,068	-0,029	0,053
		X5	-1,427	7,839	0,017	1,841	-0,403		-0,013	0,025
	2	X3	0,956	-4,619	0,081			0,107	-0,013	0,003
		X4	0,082	-1,602	0,393			-1,588	-0,089	0,065
	3	X3	0,095	-4,636	0,081			0,108	-0,013	
		X4	0,877	-2,013	0,375			-1,554	-0,084	
	X7	0,855	-6,374	-0,272			0,505	0,086		
3. adım	1	X4	-3,468	19,052	0,034	4,476		-2,074	-0,029	0,051
		X5	-1,408	7,731	0,021	1,818	-0,404		-0,013	0,025
	2	X3	0,920	-4,685	0,123			0,056	-0,009	-0,007
		X4	0,656	-1,919	0,586			-1,823	-0,069	0,022
	3	X3	0,914	-4,645	0,125			0,052	-0,009	
		X4	0,675	-2,051	0,579			-1,810	-0,067	
	X7	0,875	-6,134	-0,292			0,604	0,078		
4. adım	1	X4	-3,439	18,902	0,037	4,450		-2,076	-0,029	0,050
		X5	-1,394	7,652	0,022	1,804	-0,404		-0,013	0,025
	2	X3	0,910	-4,789	0,136			0,027	-0,006	-0,013
		X4	0,611	-2,408	0,644			-1,954	-0,054	-0,009
	3	X3	0,898	-4,710	0,140			0,019	-0,007	
		X4	0,603	-2,356	0,646			-1,960	-0,055	
	X7	0,881	-5,997	-0,298			0,663	0,077		

EK 2. Regresyon Yöntemi için İterasyonlar Sonucunda Elde Edilen Eksik Gözlem Değerleri

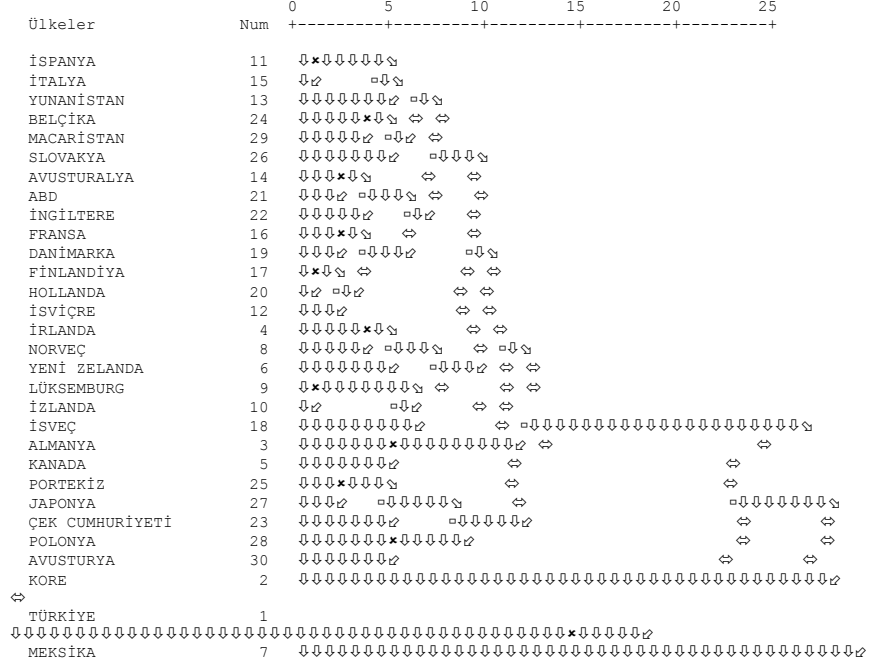
Eksik Gözlemlili Ülkeler	1.Adım				2.Adım			
	Eksik Gözleme Sahip Değişkenler				Eksik Gözleme Sahip Değişkenler			
	X3	X4	X5	X7	X3	X4	X5	X7
Belçika	0,999637	0,989077	*	*	0,997475	0,979336	*	*
Japonya	1,005177	0,98927	*	*	1,007292	0,999001	*	*
Çek Cum.	1,002751	1,029147	*	*	1,004522	1,037564	*	*
Danimarka	*	0,98619	*	*	*	0,98846	*	*
Fransa	0,997683	0,97266	*	*	1,001338	0,989499	*	*
Almanya	1,008722	1,075876	*	*	0,988431	0,98158	*	*
Yunanistan	1,009658	1,038027	*	0,723308	0,99623	0,976313	*	0,733921
İzlanda	0,988017	0,96538	*	0,769466	0,998189	1,011411	*	0,752537
İrlanda	0,985714	0,992864	*	0,774409	0,978419	0,958585	*	0,775171
İtalya	0,999775	1,039566	*	*	0,999324	1,037839	*	*
Lüksemburg	0,986985	0,984394	*	0,744798	0,995354	1,022383	*	0,731729
Hollanda	*	*	0,068496	*	*	*	0,066519	*
Norveç	*	1,000593	*	*	*	1,002118	*	*
Polonya	0,976422	0,970745	*	*	0,983682	1,004337	*	*
Portekiz	0,995294	0,983714	*	*	0,996091	0,987461	*	*
İspanya	1,006154	1,050633	*	*	1,002393	1,033518	*	*
Kore	*	*	0,064311	*	*	*	0,064631	*
Y. Zellanda	0,991164	0,975049	*	*	0,97967	0,921297	*	*
Eksik Gözlemlili Ülkeler	3.Adım				4.Adım			
	Eksik Gözleme Sahip Değişkenler				Eksik Gözleme Sahip Değişkenler			
	X3	X4	X5	X7	X3	X4	X5	X7
Belçika	0,996346	0,973989	*	*	0,995493	0,970066	*	*
Japonya	1,008514	1,004687	*	*	1,009214	1,007909	*	*
Çek Cum.	1,004394	1,036924	*	*	1,003986	1,034985	*	*
Danimarka	*	0,988945	*	*	*	0,989227	*	*
Fransa	1,003352	0,998838	*	*	1,004561	1,004429	*	*
Almanya	0,981613	0,950025	*	*	0,979193	0,939196	*	*
Yunanistan	0,991445	0,954269	*	0,736674	0,98944	0,945275	*	0,737693
İzlanda	1,002943	1,033141	*	0,74553	1,0057	1,04563	*	0,745893
İrlanda	0,977392	0,95374	*	0,771251	0,97806	0,956925	*	0,768683
İtalya	0,99789	1,031095	*	*	0,996658	1,025361	*	*
Lüksemburg	0,998929	1,038719	*	0,726342	1,000932	1,047765	*	0,723479
Hollanda	*	*	0,066553	*	*	*	0,06658	*
Norveç	*	1,002014	*	*	*	1,002164	*	*
Polonya	0,986408	1,017015	*	*	0,987666	1,022711	*	*
Portekiz	0,996566	0,989623	*	*	0,996787	0,990644	*	*
İspanya	0,999917	1,021952	*	*	0,998329	1,014638	*	*
Kore	*	*	0,065539	*	*	*	0,066432	*
Y.Zellanda	0,977423	0,91097	*	*	0,977491	0,911578	*	*

(*) Eksik gözlem içermeyen hücreler

EK 3. Ortalama ve Regresyon Yöntemleri ile Elde Edilen Veriler İçin Öklit Uzaklık Ölçüleri

Ortalama Yöntemi ile Elde Edilen Veriler İçin				Regresyon Yöntemi ile Elde Edilen Veriler İçin			
Aşama	Birleştirilen Küme		Öklit Uzaklık Katsayısı	Aşama	Birleştirilen Küme		Öklit Uzaklık Katsayısı
	Küme 1	Küme 2			Küme 1	Küme 2	
1	11	15	0,479	1	19	23	0,454
2	9	10	0,582	2	28	30	0,479
3	17	20	0,673	3	19	24	0,554
4	16	19	0,816	4	14	20	0,644
5	12	17	0,872	5	22	25	0,869
6	14	21	0,934	6	13	18	0,920
7	25	27	0,963	7	7	10	0,933
8	24	29	1,177	8	9	11	0,969
9	14	22	1,179	9	28	29	1,049
10	12	16	1,200	10	5	8	1,052
11	4	8	1,274	11	19	21	1,108
12	23	25	1,308	12	22	28	1,183
13	4	6	1,318	13	9	26	1,308
14	24	26	1,354	14	12	13	1,337
15	28	30	1,364	15	4	6	1,391
16	3	5	1,489	16	14	19	1,414
17	11	13	1,583	17	16	17	1,477
18	12	14	1,699	18	12	22	1,623
19	9	18	1,711	19	14	27	1,677
20	11	24	1,806	20	15	16	1,763
21	11	12	1,874	21	7	12	1,900
22	4	9	2,027	22	5	9	1,971
23	23	28	2,089	23	7	14	2,047
24	4	11	2,397	24	4	15	2,247
25	3	4	2,714	25	4	7	2,548
26	3	23	2,924	26	4	5	2,882
27	2	3	5,869	27	3	4	5,909
28	1	7	6,175	28	1	2	6,113
29	1	2	7,149	29	1	3	7,217

EK 4. Ortalama Yöntemi ile Elde Edilen Veriler için Ağaç Diyagramı



EK 5. Regresyon Yöntemi ile Elde Edilen Veriler için Ağaç Diyagramı

Ülkeler	Num	0	5	10	15	20	25
FRANSA	19	0					
FİNLANDİYA	23	0	0				
DANİMARKA	24	0	0				
NORVEÇ	21	0	0	0			
İZLANDA	14	0	0	0			
LÜKSEMBURG	20	0					
İSVEÇ	27	0	0	0	0		
ABD	7	0	0	0	0		
AVUSTURALYA	10	0					
SLOVAKYA	22	0	0	0			
HOLLANDA	25	0	0	0			
İTALYA	28	0	0	0			
İSPANYA	30	0	0	0			
İSVİÇRE	29	0	0	0			
BELÇİKA	13	0	0	0			
İNGİLTERE	18	0	0	0			
MACARİSTAN	12	0	0	0			
YENİ ZELANDA	4	0	0	0	0	0	0
İRLANDA	6	0	0	0	0		
ALMANYA	16	0	0	0			
KANADA	17	0	0	0			
YUNANİSTAN	15	0	0	0			
POLONYA	5	0	0	0	0		
AVUSTURYA	8	0	0	0			
PORTEKİZ	9	0	0	0			
JAPONYA	11	0	0	0			
ÇEK CUMHURİYETİ	26	0	0	0			
KORE	3	0	0	0	0	0	0
⇔							
TÜRKİYE	1	0	0	0	0	0	0
MEKSİKA	2	0	0	0	0	0	0