



Bootstrap tekniğine bağlı kümülatif toplam kontrol grafiklerinde değişim noktası analizi

Meral Yay^{1*}, Nihan Acar-Denizli², Belchin Kostov³

29.12.2015 Geliş/Received, 02.07.2016 Kabul/Accepted

doi: 10.16984/saufenbilder.54674

ÖZ

Bu çalışmanın amacı değişim noktasının belirlenmesinde kullanılmak üzere “R” programlama dili için yeni geliştirilen “ChangePoint” isimli fonksiyonun tanıtılması ve bootstrap tekniğine bağlı kümülatif toplam (CUSUM) kontrol grafikleri aracılığıyla 5 Eylül-13 Kasım 2015 tarihleri arasında İstanbul’da kaydedilen SO₂ değerlerinden oluşan veri seti üzerinde uygulanmasıdır. Değişim noktasının belirlenmesinde kestirim yöntemlerinden S_m ve MSE, grafik yöntemlerinden ise V Maskesi ve Kümülatif Toplam çizelgesi (TCUSUM) kullanılarak sonuçlar karşılaştırılacaktır.

Anahtar Kelimeler: CUSUM kontrol grafikleri, bootstrap, değişim noktası analizi, V-Mask prosedürü, TCUSUM

Change point analysis on cusum control charts based on bootstrap techniques

ABSTRACT

The aim of this study is to introduce a new function called “ChangePoint” which is developed in R language to determine change points. It is aimed to use CUSUM charts based on bootstrap technique to find out the change points of the data set consists of the recorded SO₂ values in Istanbul in the period September 5 - November 13 2015. The estimation methods S_m, MSE and the graphical methods (V Mask and Tabular Cusum) are used to determine the location of the change points and the results are compared .

Keywords: CUSUM control charts, bootstrap, change point analysis, V-mask procedure, Tabular Cusum

* Sorumlu Yazar / Corresponding Author

1 Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, İstanbul – meral.yay@msgsu.edu.tr

2 Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, İstanbul – nihan.acar@msgsu.edu.tr

3 Transverse Group for Research in Primary Care, IDIBAPS, Barcelona - badriyan@clinic.ub.es

1. GİRİŞ (INTRODUCTION)

Toplum, doğa ve insan ile ilgili tüm olaylarda karşılaşılabilecek değişimi tanımlamak istatistiksel süreç kontrolünün en önemli amacıdır. İstatistiksel yöntemler kullanılarak süreçte meydana gelebilecek değişimin neden olduğu sapmalar analiz edilebilir. W.A. Shewart (1926) tarafından ortaya atılan kontrol diyagramları, süreçte meydana gelen değişimi grafik yardımıyla açıklamaya yardımcı yöntemlerden biridir. Diyagramların en belirgin özelliği, sürecin kontrol altında olması gereken limitlerin belirlenmesinde kullanılarak ortaya çıkan değişimin gözlemlenebilmesidir. Böylece zamana bağlı olan verinin grafik üzerinde işaretlenmesi mümkün hale gelir ve belirlenen limitlerin dışında olan gözlemler için öngörülebilir. Sürecin kontrol altında olduğunun söylenebilmesi için tüm gözlemlerin alt ve üst kontrol limitleri arasında olması gerekir. Bunun sonucu olarak meydana gelen değişimin ortak sebeplere kaynaklandığı; limitler dışına çıkan gözlem(ler) mevcut ise, değişimin ortak sebeplerin yanısıra özel sebeplerden de etkilendiği söylenebilir.

Sürecin kontrolünde gözlemlerin yapısına bağlı olarak farklı türde kontrol diyagramları geliştirilmiştir. Shewhart ile başlayan sınıflandırma, 1954 yılında E.S. Page tarafından geliştirilen CUSUM ile devam etmiştir. Ağırlıklı hareketli ortalamayı da diyagram içinde analize imkan sağlayan EWMA ise ilk kez 1959 yılında S. W. Roberts tarafından ortaya atılmıştır. Sürecin belirli olasılık dağılımları için yapılan varsayımları sağlaması durumunda kullanılan dağılıma bağlı grafikleri ilk kez 2001 yılında D.C. Montgomery önerirken, herhangi bir dağılım varsayımı gerektirmeyen kontrol grafiklerini ise 2001 yılında P. Van der Laan and S.T. Bakir ve 2002 yılında S.T. Bakir tarafından ortaya konularak literatürdeki yerini almıştır. Sayılan kontrol diyagramlarının çoğu sürecin belirli bir olasılık dağılımına uyduğu varsayımı altında çalışır. Dağılım şeklinin sürekli ölçüm değerleri için normal, nitel veri için Binom ya da Poisson dağılımına uyması beklenir. Ancak uygulamada varsayımların her zaman sağlanmadığı sık karşılaşılan bir durumdur. Buna bağlı olarak parametrik varsayımların sağlanmadığı durumlarda parametrik olmayan bir başka deyişle dağılım ile ilgili varsayım gerektirmeyen yöntemler kullanılmaktadır.

CUSUM Kontrol Diyagramları, düzenli aralıklarla yapılan ölçümlerin önceden belirlenen değerlerle karşılaştırıldığı araştırmalarda kullanılan bir yöntemdir [1]. Bu diyagramların en önemli özelliği, süreçte meydana gelen küçük değişkenlikleri tespit edebilmesidir. Ayrıca değişimin hangi gözlemden

itibaren başladığını belirlemede oldukça başarılıdır ve süreçteki sapmayı diğer kontrol grafiklerinden daha büyük hassasiyetle ortaya koymaktadır. Gözlemlenen kaymanın yerini belirlemek kaymanın sebebinin belirlemede önemli bir rol oynar [2]. Kümülatif toplam kontrol grafikleri, sabit bir sayıdan seride yer alan değerlerin her birinin çıkarılması ve elde edilen ilk farka, sonraki farkların eklenmesiyle oluşan kümülatif toplam değerlerin koordinat sisteminin absis eksenine yerleştirilmesiyle elde edilir. Çizilen grafik yardımıyla eğimde meydana gelen değişime bakılarak sürecin ortalamasındaki kayma rahatlıkla gözlemlenebilir. Ortalamada meydana gelen ani değişiklikleri göstermede başarılı olan CUSUM diyagramlarına bakılarak, düz çizgi şeklinde elde edilen yolun ortalamada değişiklik olmayan bir periyodu işaret ettiği söylenebilir. Ortalamadan kayma meydana gelen noktalarda ise eğim oluşur [3]. Eğer değişiklik pozitif yönde ise pozitif eğim, negatif yönde ise negatif eğim meydana gelir [4]. CUSUM diyagramlarında kontrol dışı noktaların belirlenmesi için V maskesi veya Kümülatif Toplam Çizelgesi (TCUSUM) grafiksel yöntemlerinin yanısıra Bootstrap tekniğine dayalı kestirim yöntemleri de kullanılmaktadır.

Bu çalışmada "R" programlama dili için bahsedilen yöntemleri bir arada kullanabilen bir fonksiyon geliştirilmiştir. Bu fonksiyon 5 Eylül - 13 Kasım 2015 tarihleri arasında İstanbul'da kaydedilen SO₂ değerlerinden oluşan veri setine ait değişim noktalarının tespit edilmesinde kullanılmıştır. Değişim noktasının belirlenmesinde kestirim ve grafik yöntemleri birlikte ele alınarak, sonuçlar karşılaştırmalı olarak sunulmuştur.

1.1. Bootstrap Tekniğinin CUSUM Kontrol Diyagramlarına Uyarlanması (The Adaptation of Bootstrap Technique on CUSUM Control Charts)

Parametrik varsayımların sağlanmadığı ve dağılım şekli hakkında önbilginin olmadığı durumda kontrol limitlerini elde etmek için çok sayıda bootstrap tekrarlı örnekleme elde edilebilir. Tekrarlı örneklemlerin sayısının çokluğu, günümüzde kullanılan bilgisayar programları yardımıyla sorun olmaktan çıkmıştır. Bunun yanısıra geliştirilen programlar yardımıyla daha hızlı ve düşük maliyetli bir uygulama yapma imkanı vardır [5].

CUSUM kontrol diyagramlarının elde edilebilmesi için veriye bağlı olarak kümülatif toplamların hesaplanması ve elde edilen değerlerin grafiğe dönüştürülmesi gerekir. X_1, X_2, \dots, X_n "n" adet veriyi ifade etmek üzere, $S_0, S_1, S_2, \dots, S_n$ ile gösterilen kümülatif toplamların elde edilmesi için ilk olarak $\bar{X} = (X_1 + X_2 + \dots + X_n) / n$ ile ortalama hesaplanır.

Süreç ile ilgili önbilgi varsa " \bar{X} " yerine, sürecin beklenen değerini ifade eden hedef değer μ_0 kullanılabilir [6]. Kümülatif toplamları oluşturmak için $S_0 = 0$ kabul edilerek S_i değerleri, $S_i = S_{i-1} + (X_i - \bar{X})$, $i = 1, 2, \dots, n$ iteratif olarak elde edilir. Kümülatif toplam değerlerin zaman içinde ortalama etrafında toplanma eğilimi gösterdiği varsayılır. Değerlerin toplanmasıyla elde edilen bir seri olduğu için pozitif artan bir yapıya sahiptir. S_{diff} değeri, $S_{diff} = S_{max} - S_{min}$ olarak belirlendikten sonra kontrol limitlerine ulaşmaya çalışılır ve bir tekrar süreci içine girilir. Burada $S_{max} = \max S_i$ ve $S_{min} = \min S_i$ olarak kabul edilir.

Bootstrap tekniğinin CUSUM kontrol diyagramlarına uyarlanabilmesi için öncelikle $X_1^0, X_2^0, \dots, X_n^0$ bootstrap örneklemi yaratılır. İadesiz seçim yöntemiyle yaratılan bu örneklemeye dayanarak bootstrap CUSUM değerleri $S_0^0, S_1^0, S_2^0, \dots, S_n^0$ elde edilir. Buradan hareketle kümülatif toplamlar içindeki min, max ve fark değerleri S_{min}^0, S_{max}^0 ve S_{diff}^0 elde edilerek orjinal fark değeri olan S_{diff} ile karşılaştırılır.

Bu süreç baştan itibaren "N" kez tekrarlanarak simülasyon çalışması tamamlanır. Bootstrap simülasyonunun temelinde yatan düşünce, bootstrap örneklemelerinin veriye ilişkin tekrarlı sıralanışları ifade etmesidir. Böylece çok sayıda bootstrap örneği yaratılarak S_{diff} değerinin ne kadar değişime uğradığını tespit etmek mümkün hale gelir. Bootstrap örneklemelerinin sayısı arttıkça daha iyi tahminler yapılabilir. "I", $S_{diff}^0 < S_{diff}$ eşitsizliğini sağlayan bootstrap örneklemelerinin sayısını göstermek üzere değişimin meydana geldiği noktadaki güven düzeyi (confidence level) aşağıdaki gibi elde edilir [7].

$$CL = 100 \frac{I}{N} \% \quad (1)$$

Bootstrap CUSUM tekniğinin uygulanabilmesi için bağımsız hata yapısı varsayımının sağlanması gerekir. Hem kontrol diyagramının hem de değişim noktası analizinin bootstrap yapısında uygulanabilmesi için ortalama değişim modeline uyması gerekir. X_1, X_2, \dots, X_n zamana bağlı gözlem değerlerini ifade etmek üzere ortalama değişim modeli, $X_i = m_i + e_i$ olarak ifade edilir. Burada " m_i ", " i " zamanındaki ortalama değer ve " e_i ", " i " değere ait rasgele hatayı ifade eder. Hataların ortalaması sıfırdır ve birbirlerinden bağımsızlardır.

Örneklem sayısı birden fazla olduğunda ortalamalar yerine standart sapmalar üzerinden Bootstrap CUSUM uygulanarak değişkenlikteki değişim analiz edilebilir [8]. Bahsedilen yöntemin yanısıra Chatterjee ve Qiu (2009) yine bootstrap yöntemine dayalı olarak CUSUM çizelgelerinde kontrol dışı gözlemlerin belirlenmesi için dağılımdan bağımsız ve sağlam bir yöntem önermişlerdir [9].

2. DEĞİŞİM NOKTASINI BELİRLEME YÖNTEMLERİ (METHODS FOR DETECTING CHANGE POINT)

Değişim noktası analizinin amacı, sürecin ortalamasında meydana gelen herhangi bir "değişim" olup olmadığını belirlemektir. Değişim analizindeki temel varsayım gözlemlerin bağımsız olması ve aynı dağılıma sahip olmalarıdır. Literatürde gözlemler arasında güçlü bir otokorelasyonun varlığı halinde değişim noktası analizi için CUSUM diyagramını kullanan uygulamalara rastlanabilir [10].

2.1. Değişim Noktasını Belirlemede Kestirim Yöntemleri: S_m ve MSE Kestiricileri (The Estimation Methods Used for Detecting Change Point: S_m and MSE Estimators)

Değişim noktasının belirlenmesinde kullanılan S_m kestiricisi $|S_m| = \max_{i=0,1,2,\dots,n} |S_i|$ olarak tanımlanır. CUSUM diyagramında "0" dan en uzak nokta olan "S" noktası, değişimin meydana geldiği noktadan bir önceki nokta olan "m" yardımıyla tanımlanır. Değişimin belirlenmesinde kullanılan diğer tahmin edici ise ortalama kare hata (mean square error, MSE) dir. Bu tahmin edicide veri seti öncelikle 1 den m ' e ve $m+1$ ' den n ' ye kadar olmak üzere iki parçaya ayrılır. MSE tahmin edicisi ile belirlenmek istenen, hangi parçanın öngörülen ortalamaya uygun olduğunu belirlemektir. "m" noktasının MSE kestiricisi,

$$MSE(m) = \sum_{i=1}^m (x_i - \bar{x}_1)^2 + \sum_{i=m+1}^n (x_i - \bar{x}_2)^2 \quad (2)$$

olarak tanımlanır. MSE eşitliğinde kullanılan \bar{x}_1 ve \bar{x}_2 sırasıyla aşağıdaki gibidir [7]:

$$\bar{x}_1 = \frac{\sum_{i=1}^m x_i}{m} \quad \text{ve} \quad \bar{x}_2 = \frac{\sum_{i=m+1}^n x_i}{n-m} \quad (3)$$

Eşitlikteki "m" noktası MSE'yi minimize eden değerdir ve değişimden önceki son noktayı veren en iyi tahmin edicidir. "m+1" ise değişimin başladığı noktadır.

Böylece değişim noktası tanımlanarak, bootstrap yönteminin uygulanması ile yukarıda tanımlanan sürecin güven aralıklarının elde edilmesi mümkündür.

2.2. Değişim Noktasını Belirlemede Grafikselleştirme Yöntemleri: V Maskesi ve Kümülatif Toplam Çizimleri (The Graphical Methods Used for Detecting Change Point: V Mask and TCUSUM Charts)

V maskesi, kontrol dışı noktaların tespit edilmesi amacıyla kümülatif toplam grafiklerine tamamlayıcı olarak kullanılan bir yöntemdir. CUSUM grafiğindeki kümülatif toplam değerleri, V maskesinin kolları arasında kalıyorsa sürecin kontrol altında olduğu, eğer V maskesinin kollarının dışına kayan değerler varsa sürecin kontrol dışına çıktığı söylenebilir [11].

V maskesi çizimi için temelde iki yöntem vardır. Birinci yöntemde V maskesi 1. Tip hata α , 2. Tip hata β ve ortalamadaki kayma miktarı olan δ parametrelerine bağlı olarak çizilir. Burada α "bir kayma yokken kayma tespit edilmesi olasılığı", β ise "gerçekte kayma varken tespit edilmeme olasılığı" dır. Bu parametreler kullanılarak V maskesinin üzerine yerleştirildiği gözlem ile köşe noktası arasındaki mesafeyi ifade eden " d " parametresi aşağıdaki gibi bulunur:

$$d = \frac{2}{\delta^2} \ln\left(\frac{1-\beta}{\alpha}\right) \quad (4)$$

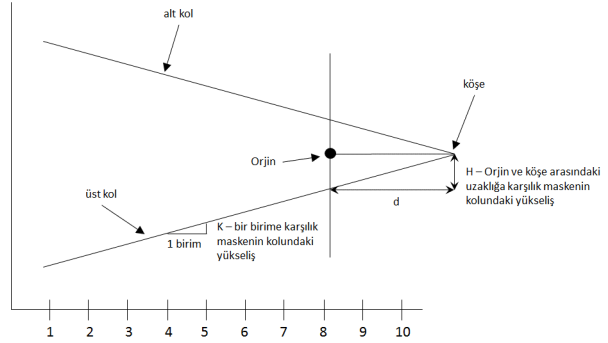
β değeri çok küçük olduğunda bu uzaklık α parametresine bağlı olarak hesaplanır.

$$d = \frac{-2}{\delta^2} \ln \alpha \quad (5)$$

V maskesinin kollarının eğimi, maskenin köşe noktası ile üst ya da alt kolu arasında kalan açısının tanjantına eşittir ve $\tan \theta$ olarak ifade edilir.

İkinci yöntemde ise V maskesi, karar aralığının bir ölçüsü olan h ve referans değeri olarak bilinen k parametrelerine bağlı olarak çizilir. k değeri genellikle $\delta/2$ ve h değeri ise 4 ya da 5 olarak alınır [6, 12]. Bu parametrelerin örneklem standart hatası $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ ile çarpımıyla karar aralığı $H = h\sigma_{\bar{x}}$ ve V maskesinin eğimi olan $K = k\sigma_{\bar{x}}$ olarak hesaplanır. Tek bir örneklem söz konusu ise $n=1$ olarak alınır ve bu durumda standart hata standart sapmaya eşit olur. Bu yöntem hesap kolaylığı sağladığı için birinci yöntem tercih edilir. Örnek bir V maskesi çizimi Şekil 1.'deki

gibidir. Yerli literatürde V maskesi çizimi konusunda ayrıntılı bilgi Oktay ve Özçomak (2001) tarafından yapılan çalışmada verilmiştir.



Şekil 1. Örnek V maskesi çizimi (Example of Vmask graph)

V maskesi veri setine eklenen veya maskenin kolları dışına taşan her bir gözlem için tekrar çizilmesi gerektiğinden kullanımı zor bir yöntemdir. İki yönlü bir yapısı olduğundan tek yönlü süreçlerde kullanılması tavsiye edilmez. Buna rağmen değişim noktalarının grafikselleştirme yöntemiyle belirlenmesinde ve sürecin güven aralığının çizilmesinde V maskesi kullanımı yaygındır. Hedef değer üstünde ve altında kalan ortalamadan sapmaların belirlenmesinde V maskesine alternatif olarak tek yönlü değişimi araştırmak için Kümülatif Toplam Çizelgesi (TCUSUM) yöntemi de kullanılabilir. Başlangıç değerleri $S_{hi} = S_{li} = 0$ olmak üzere, ham veri için ortalamanın üstünde ve ortalamanın altında kalan gözlemleri belirlemek için üst sınır ve alt sınır kümülatif toplamaları sırasıyla aşağıdaki eşitliklerden hesaplanır:

$$S_{hi} = \max[0, x_i - (\mu_0 - k) + S_{h(i-1)}] \quad (6)$$

$$S_{li} = \max[0, (\mu_0 - k) - x_i + S_{l(i-1)}] \quad (7)$$

Standardize edilmiş $z_i = (x_i - \mu_0)/\sigma$ değerleri için ise bu değerler (8) ve (9) no'lu eşitliklerden bulunur.

$$S_{hi} = \max[0, z_i - k + S_{h(i-1)}] \quad (8)$$

$$S_{li} = \max[0, -k - z_i + S_{l(i-1)}] \quad (9)$$

Bu eşitliklerden elde edilen S_{hi} ve S_{li} değerleri, veriden hesaplanan H değeri ile karşılaştırılarak yorum yapılır. H değerini ilk aşan değere ait gözlem kontrol dışı bir sürecin başladığı noktayı gösterir.

3. PROGRAMLAMA (PROGRAMMING)

Literatürde CUSUM grafiklerini uygulayan farklı programlar ve fonksiyonlar bulunmaktadır [13-16]. Taylor (2014), bootstrap tekniğine bağlı CUSUM kontrol grafikleri yardımıyla değişim noktasının bulunmasını amaçlayan "Change Point Analyzer" isimli programı geliştirmiştir. V maskesi ve TCUSUM grafiksel yöntemleri SAS'a ait olan JMP programında mevcuttur. Bunun dışında R programındaki "qcc" paketinde tek yönlü değişimlerin araştırılması için TCUSUM çizen "cusum" fonksiyonu bulunmaktadır. Bu yöntemleri ayrı ayrı yapan programlar olmasına rağmen tüm yöntemleri bir arada uygulayan bir programa rastlanmamıştır. Bu çalışmada, R programında bootstrap tekniğine bağlı kestirim yöntemleri ile grafiksel yöntemleri aynı anda yapmaya olanak sağlayan "ChangePoint" isimli yeni bir fonksiyon geliştirilmiştir. Bu fonksiyona ait giriş parametre değerleri şu şekildedir:

```
ChangePoint(data , nb.boots = 1000, conf.lim = 0.95,
level = 3, n = 1, sigma = 3, beta = 0.01, delta = 1, h = 4,
pos.vmask = length(data), method = 1, aver = NULL,
st.dev = NULL)
```

Fonksiyona ait giriş parametrelerin neyi ifade ettikleri aşağıdaki gibi özetlenebilir:

- data: Değişim noktası analizi yapılacak veriyi içeren vektör.
- nb.boots: Bootstrap simülasyon sayısı.
- conf.lim: Anlamli değişim ile ilişkili güven düzeyi.
- level: Değişim tespit etmek için yapılan aşama sayısı.
- n: Örneklem sayısı.
- sigma: 1. Tip hata α 'yı tespit etmek için kullanılan kriter.
- beta: Proses ortalamasında meydana gelen değişikliğin tespit edilmemesi olasılığı.
- delta: Ortalamada tespit edilmesi beklenen kayma miktarı.
- h: V maskesinin kolunun orjinden dikey uzaklığını belirlemede kullanılan parametre.
- pos.vmask: V maskesinin üzerine yerleştirildiği nokta.
- method: V maskesinin çiziminde kullanılan yöntem. Method "1" h ve k parametrelerine bağlı, Method "2" sigma ve beta parametrelerine bağlı olarak V maskesini çizdirir.
- aver: Kullanıcı tarafından girilen hedef değer. Sürece ait hedef değer yoksa sürecin teorik ortalaması olarak alınır.

- st.dev: Prosese ait standart sapma değeri, belirtilmediyse sürecin teorik standart sapması olarak alınır.

Fonksiyondan elde edilen çıktı parametreleri ise şunları ifade etmektedir:

- change.points: 1. Satırda değişim noktasını, 2. Satırda bu noktaya ait güven düzeyini gösteren matrisleri içeren bir liste
- op.cpts: Belirtilen güven düzeyindeki değişim noktalarının sayısı.
- cusum.tabular: Kümülatif Toplamlar Çizelgesi (TCUSUM).
- H: Kontrol dışı noktaları belirlemek için kullanılan limit değeri.

"ChangePoint" fonksiyonu ile altı adet grafik elde edilir. Bunlar veri seti için 2σ ve 3σ güven aralıklı kontrol tablosu, CUSUM kontrol grafiği, Bootstrap örnekleminden elde edilen S_{diff} değerlerinin dağılımını gösteren histogram grafiği, ilk 5 bootstrap örneğine ait CUSUM çizimleri, V mask ile CUSUM grafiği ve ortalamadaki azalış ve artışı gösteren TCUSUM grafikleridir.

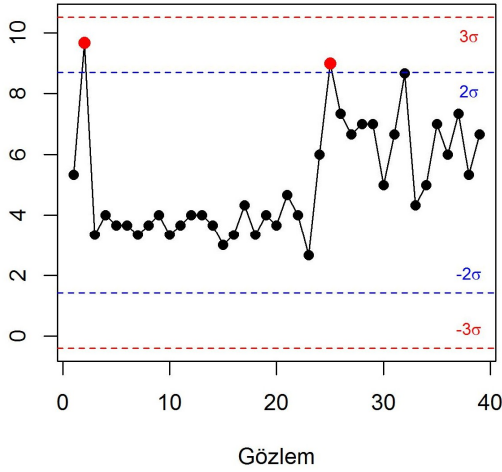
"ChangePoint" fonksiyonun diğer CUSUM grafiklerini uygulayan programlara ve fonksiyonlara oranla iki önemli avantajı vardır. İlki parametrelerin kullanıcı tarafından kolaylıkla değiştirilebilmesidir. Birinciden daha önemli olan bir diğeri ise, fonksiyonun kodunun açık bir kod olmasıdır. Böylece, daha ileri derecede programlama bilgisi olan kullanıcıların kodu ihtiyaçları doğrultusunda kolaylıkla değiştirerek geliştirebilmesi mümkündür.

4. UYGULAMA (APPLICATION)

Bu çalışmada İstanbul'da 5 Eylül - 13 Ekim 2015 tarihleri arasında kaydedilmiş havadaki kükürt dioksit (SO_2) değerlerinin uğradığı değişimi analiz etmek ve varsa değişim noktalarını belirlemek amacıyla bootstrap tekniğine bağlı CUSUM kontrol grafikleri kullanılmıştır. Bu bağlamda, Ulusal Hava kalitesi izleme ağı web sitesinden (<http://www.havaizleme.gov.tr/Default.ltr.aspx>) İstanbul'un Aksaray, Beşiktaş ve Ümraniye ilçelerinde bulunan 3 istasyona ait SO_2 değerlerinin günlük ortalamaları alınmıştır. Veri seti toplam 39 gözlemden oluşmakta ve her gün için 3 istasyondan alınan tekrarlı değerleri içermektedir.

Kullanılan veri setinde Beşiktaş istasyonuna ait 20 numaralı değer gözlenememiş, bu nedenle 23 Eylül 2015 tarihindeki ortalama SO_2 değeri kullanılmıştır. 5 Eylül -

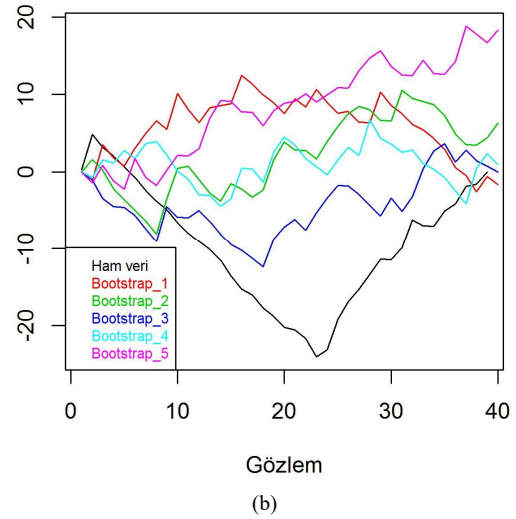
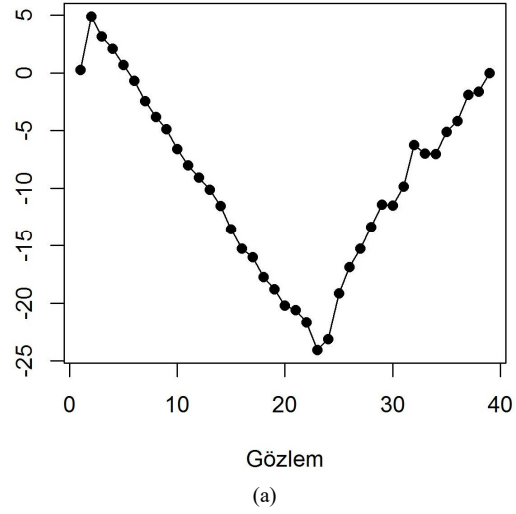
13 Ekim 2015 periyodunda 3 istasyona ait günlük ortalama (SO_2) değerleri Şekil 2'deki gibidir.



Şekil 2. 2σ ve 3σ güven aralıklı Shewart tablosu (Shewart table with 2σ and 3σ confidence interval)

2σ ve 3σ güven aralıkları ile Shewhart kontrol grafiği oluşturulduğunda 3σ güven aralığı için kontrol limitleri dışına çıkan bir gözlem bulunmamıştır. 2σ güven limiti için ise sürecin başlangıcında ve ortasında birer gözlemin kontrol limiti dışına çıktığı görülmektedir. Ancak Shewhart kontrol grafiği, hem normallik varsayımı gerektirdiği hem de küçük oranda sapmaları tespit etmede yetersiz kaldığı için kullandığımız veri setinde değişim noktası belirlemek için uygun değildir. Bundan dolayı, hem dağılım için varsayım gerektirmediğinden hem de ortalamadan küçük oranda sapmaları bulmada daha etkin olduğundan dolayı CUSUM çizelgelerinin bootstrap yöntemi ile birlikte kullanılmasına karar verilmiştir. Varsayımları test etmek için Taylor (2014) tarafından geliştirilen "Change Point Analyzer 2.3" programı kullanılmıştır. Veri setinde aykırı değer olmadığı ve hataların bağımsızlığı varsayımının sağlandığı görülmüştür. Bu durumda, CUSUM tabloları ve değişim noktası analizi güvenle uygulanabilir.

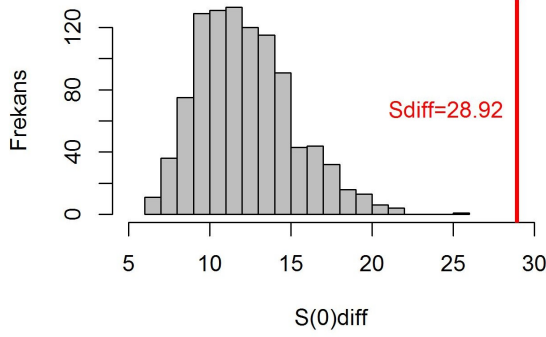
Üç istasyondan alınan verilerin ortalamalarına ait CUSUM çizelgesi Şekil 3a'daki gibidir. Grafikten görüldüğü gibi belli bir noktaya kadar CUSUM değerleri negatif eğim gösterirken bir noktadan itibaren pozitif eğim göstermeye başlamıştır. Bu durum gözlenen SO_2 değerlerinin ilk periyotta genel ortalamasının altında kaldığı, ancak değişim noktasından sonra ortalamasının üstüne çıktığı anlamına gelmektedir.



Şekil 3. a) Kümülatif Toplam Kontrol Grafiği (CUSUM Control Chart), b) İlk 5 Bootstrap örneklemleri (First 5 Bootstrap samples)

Kümülatif Toplam değerleri üzerinden ilk 5 iadesiz bootstrap örneklemlerine ait dağılım çizildiğinde orijinal CUSUM grafiğinden farklı dağılım göstermektedir (Şekil 3b). Bu durum veri seti içinde bir değişim noktası olabileceğine işaret etmektedir.

Değişim noktası belirlenebilmesi için CUSUM değerleri üzerinden iadesiz 1000 adet bootstrap örneklemleri oluşturulmuştur. Elde edilen 1000 adet örneklemin maksimum ve minimum CUSUM değerleri arasındaki fark alınarak asıl örneklemden elde edilen S_{diff} değeri ile karşılaştırılmıştır. Bootstrap örneklemlerinden edilen S_{diff} değerlerine ait histogram Şekil 4'de verildiği gibidir. Bu grafiğe göre orijinal S_{diff} değerinin üzerine çıkan bir değer bulunamamıştır. Bu durum veri setinde %100 güvenle bir değişim noktası olduğunu işaret eder.



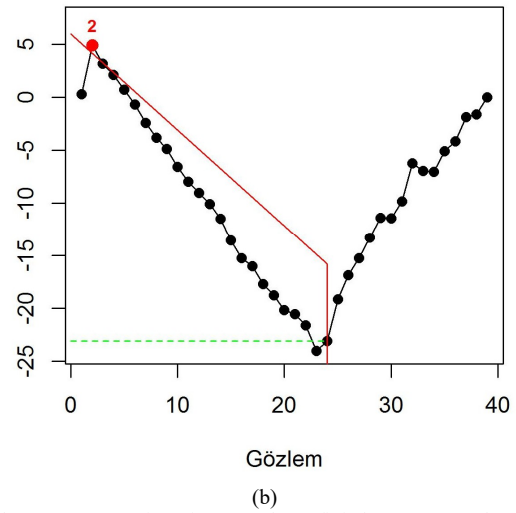
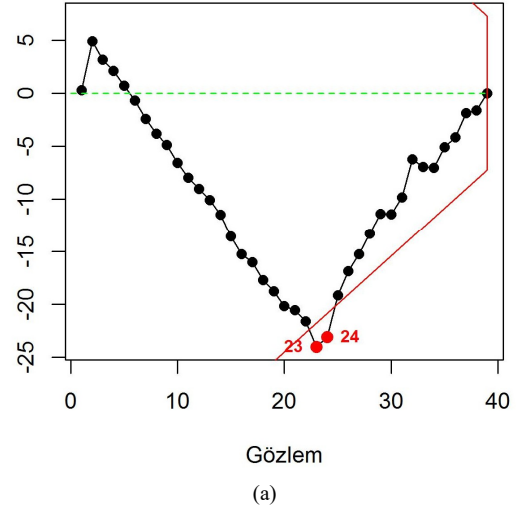
Şekil 4. S_{diff} değerlerinin dağılımı (The histogram of S_{diff} values)

Değişim noktasının yerini tespit etmek için Bootstrap CUSUM yöntemi kullanılmıştır. MSE değerlerine bakılarak değişimden önceki son noktanın 23 numaralı gözlem olduğu görülmüştür (Tablo1). Dolayısıyla, 24 numaralı gözlemden itibaren S_{O_2} değerlerinde değişim olduğu anlamına gelmektedir. 28 Eylül 2015 tarihine denk gelen bu gün hem okulların açıldığı hem de 9 günlük Ramazan Bayramı tatilini takip eden ilk iş günüdür. Bu bağlamda, bu noktadan itibaren görülen S_{O_2} değerlerinin ortalamasındaki artışın trafiğe çıkan araç sayısından kaynaklandığı söylenebilir. İkinci ve üçüncü aşamalarda yüksek güven düzeyiyle değişim noktası olduğu düşünülen bir gözleme rastlanmamıştır.

Tablo 1. S_{O_2} değerleri için değişim noktası analizi sonuçları (Results of Change Point Analysis for S_{O_2} values)

| Gözlem | Tarih | Güven düzeyi | Aşama |
|--------|------------|--------------|-------|
| 24 | 28.09.2015 | 100.0% | 1 |
| 3 | 07.09.2015 | 72.1% | 2 |
| 33 | 07.10.2015 | 45.7% | 2 |
| 2 | 06.09.2015 | 48.6% | 3 |
| 23 | 27.09.2015 | 48.9% | 3 |
| 32 | 06.10.2015 | 47.5% | 3 |
| 35 | 09.10.2015 | 73.3% | 3 |

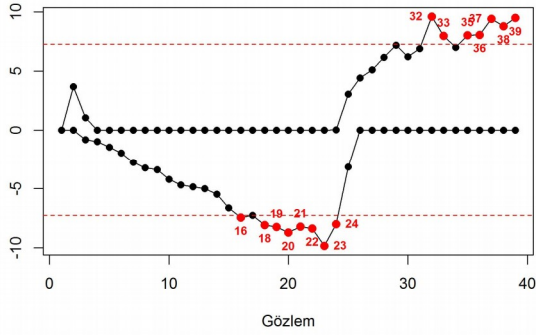
Grafiksel olarak değişim noktalarının tespit edilmesi için V maskesi incelenmiştir. Veri setine ait V maskesi kayma miktarı $\delta = 1$, $h = 4$ ve $k = 1/2$ alınarak Şekil 5a'daki gibi elde edilmiştir. V maskesinin alt sınırını aşan iki gözlem bulunmuştur. Son gözlemden geriye doğru gelindiğinde maske dışına taşan ilk noktanın 24 numaralı gözlem olduğu görülür. V maskesi 24 numaralı gözlem üzerine yerleştirilerek tekrar çizildiğinde 2 numaralı gözlemin V maskesinin üst kolu dışına taşıdığı gözlenir (Şekil 5b). Bu durumda, veri setindeki 23 ve 24 numaralı gözlemlerin yanısıra 2 numaralı gözlem de kontrol dışı nokta olarak tespit edilmiştir.



Şekil 5. a) Ham veriye ait V mask grafiği, b) 24 numaralı gözlem üzerine yerleştirilerek çizilen V mask grafiği

Tek yönlü değişimleri incelemek için TCUSUM çizildiğinde, alt sınırdaki 16. gözlemin ve 18 ile 24 numaralı gözlemler arasındaki değerlerin ortalamasının altında kaldığı görülmüştür (Şekil 6). Üst sınırdaki ise 32. gözlemden itibaren 34 numaralı gözlem hariç tüm gözlemlerin ortalamasının üstünde değerler aldığı söylenebilir.

Grafiksel yöntemleri karşılaştırdığımızda çift yönlü yapısı olan V maskesi TCUSUM'dan daha az sayıda kontrol dışı nokta tespit etmiştir. Bootstrap CUSUM ile değişim noktası olarak tespit edilen 24 numaralı gözlemin her iki grafiksel yöntemde de kontrol dışı noktalar arasında yer aldığı görülmüştür. Ancak $\max|S_m|$ ve MSE kestirim yöntemleri sonucunda 24 numaralı gözlem dışında yüksek güven düzeyinde başka bir değişim noktası görülmemiştir.



Şekil 6. Kümülatif Toplam Çizelgesi (Tabular Cusum)

5. SONUÇ (CONCLUSIONS)

5 Eylül - 13 Ekim 2015 tarihleri arasında havadaki SO₂ değerlerinin uğradığı değişimin analizi R programında geliştirilen yeni bir fonksiyon yardımıyla yapılmıştır. Bootstrap CUSUM ile yüksek güven düzeyinde bir değişim noktasının var olduğu tespit edilmiştir. Grafikselleştirilen yöntemlerden V mask ve TCUSUM ile kontrol dışı gözlemler belirlenmiş ancak hangi noktanın değişim noktası olduğunu belirlemede $\max|S_m|$ ve MSE kestirim yöntemlerinden faydalanılmıştır. Değişim noktası olarak tespit edilen 28 Eylül 2015 tarihli 24 numaralı gözlem tam da beklendiği gibi havadaki SO₂ değerinin ortalamasının üstünde seyretmeye başladığı güne denk gelmiştir. Geliştirilen fonksiyon yardımıyla, tüm bu analizleri tek bir program altında yapmak mümkündür. Kullanıcıların ücretsiz olarak faydalanabilecekleri bu fonksiyon makale yazarlarından talep edilebilir.

6. KAYNAKÇA (REFERENCES)

- [1] R. H. Woodward ve P. L. Goldsmith, Cumulative Sum Techniques: Mathematical and Statistical Techniques for industry, *ICI Monograph 3*, 1964.
- [2] R. H. Lester ve N. L. Encrick, Quality Control for Profit, New York: Marcel Dekker Inc., 1992.
- [3] P. Gavit, Y. Baddour ve R. Tholmer, "Use of change-point analysis for process monitoring and control," *BioPharm International*, cilt 22, no. 8, 2009.
- [4] T. Staphenurst, Mastering Statistical Process Control: A Handbook for Performance Improvement Using Cases, UK: Elsevier, 2006.
- [5] S. Teyarachakul, S. Chand ve J. Tang, "Estimating the Limits for Statistical Process Control Charts: A Direct Method Improving upon the Bootstrap," *European Journal of Operational Research*, cilt 178, no. 2, pp. 472-481, 2006.
- [6] G. Henderson, D. R. Robin ve D. MacDonald, "Bringing Data to Life with Post-Hoc CUSUM Charts," *Case Studies In Business, Industry and Government Statistics*, cilt 3, no. 2, pp. 60-69, 2010.
- [7] A. Khan, S. Chatterjee, D. Bisai ve N. K. Barman, "Analysis of Change Point in Surface Temperature Time Series Using Cumulative Sum Chart and Bootstrapping for Asansol Weather Observation Station, West Bengal, India," *American Journal of Climate Change*, cilt 3, pp. 83-94, 2013.
- [8] W. A. Taylor, "Change Point Analysis: A Powerful New Tool For Detecting Changes," [Çevrimiçi]. Available: <http://www.variation.com/cpa/tech/changepoint.html>. [Erişildi: 26 Mayıs 2014].
- [9] S. Chatterjee ve Q. Peihua, "Distribution-Free Cumulative Sum Control Charts Using Bootstrap-Based Control Limits," *The Annals of Applied Statistics*, cilt 3, no. 1, pp. 349-369, 2009.
- [10] O. O. Atienza ve B. W. Ang, "A CUSUM Scheme for autocorrelated observations," *Journal of Quality Technology*, cilt 34, no. 2, pp. 187-199, 2002.
- [11] E. Oktay ve M. S. Özçomak, "CUSUM Kontrol Grafikleri ve Bir Uygulama," *İktisadi ve İdari Bilimler Dergisi*, cilt 15, no. 1-2, pp. 361-387, 2001.
- [12] V. V. Koshti, "Cumulative Sum Control Chart," *International Journal of Physics and Mathematical Sciences*, cilt 1, no. 1, pp. 28-32, 2011.
- [13] L. Scrucca, "qcc: an R package for quality control charting and statistical process control," *R News*, cilt 4, no. 1, pp. 11-17, 2004.
- [14] W. Taylor, "Change Point Analyzer 2.3 Software Package," Taylor Enterprises Inc., Libertyville, 2014.
- [15] JMP 12 Online Documentation , "The Cusum Control Chart," [Çevrimiçi]. Available: http://www.jmp.com/support/help/The_CUSUM_Control_Chart.shtml. [Erişildi: 25 Kasım 2015].
- [16] S. Knoth, "spc: Statistical Process Control - Collection of Some Usefull Functions," R package version 0.5.3, 08 02 2016. [Çevrimiçi]. Available: <https://cran.rproject.org/web/packages/spc/index.html>. [Erişildi: 18 06 2016].