

# Potasyum (K) ve Hemoliz İndeksi (HI) Arasındaki İlişkinin Kümeleme Analizi Yöntemi ile İrdelenmesi

Kemal TURHAN<sup>a</sup>, Yasemin Zeynep ENGİN<sup>a</sup>, Sabiha KAMBUROĞLU<sup>b</sup>, Burçin KURT<sup>a</sup>, Asım ÖREM<sup>b</sup>

<sup>a</sup> Biyoistatistik ve Tıp Bilişimi AD, Karadeniz Teknik Üniversitesi, Trabzon

<sup>b</sup> Tıbbi Biyokimya AD, Karadeniz Teknik Üniversitesi, Trabzon

## Analysis of the Relation between Potassium (K) and Haemolysis Index (HI) Using Clustering Method

**Abstract:** In this study, the relation between Potassium (K) and Haemolysis Index (HI) was analysed with Clustering method using data which were derived from patients hospitalized between June 2011-June2012 at Karadeniz Technical University Faculty of Medicine Farabi Hospital. As a result of pre-processing step, number of records was reduced to 1252. Records were consist of last two K and HI measurements and absolute differences, measurement dates as data columns. Correlation between the absolute value of the differences was analysed using K and HI results in five clusters. Clusters were created using k-Means algorithm and Euclidian distance measurement equation with the help of Statistica package program. The results which had HI difference values higher than 295 mg/dL were showed a very high correlation. Cluster analysis can be used as a preliminary study for correction factor studies to prevent repetition of the tests.

**Key Words:** Haemolysis, Potassium, Correlation, Clustering

**Özet:** Bu çalışmada Haziran 2011-Haziran 2012 tarihleri arasında Karadeniz Teknik Üniversitesi Tıp Fakültesi Farabi Hastanesi'nde yatan 9777 hastanın verileri kullanılarak potasyum (K) ve hemoliz indeksi (HI) arasındaki ilişki veri madenciliği tekniklerinden Kümeleme Analizi (Clustering) yöntemi kullanılarak irdelenmiştir. Analiz öncesinde yapılan ön işlemler sonucunda kayıt sayısı 1252'ye düşürülmüştür. Analiz için süzülen kayıtlarda alınan son iki K ve HI sonuçları, sonuçların kayıt tarihleri, ilk ve son K ve HI değerlerinin farklarının mutlak değerleri alınarak oluşturulan farkları bulunmaktadır. Analiz sonucunda, K ve HI farklarının mutlak değerleri kullanılarak bulunan kümelerde mutlak K ve HI farkları arasındaki korelasyon ölçülmüştür. Kümeler k-Means algoritması ve Euclidian mesafe ölçüm tekniği kullanılarak Statistica yazılım paketi kullanılarak oluşturulmuştur. Bulunan sonuçlar HI farkının 295 mg/dL' den yüksek olduğu kümelerde çok yüksek korelasyon olduğunu göstermektedir. Hemolize örneklerde test tekrarı engel olacak düzeltme faktörü çalışmaları için kümeleme analizinin ön çalışma olarak kullanılabilceği sonucuna varılmıştır.

**Anahtar Kelimeler:** Hemoliz, Potasyum, Korelasyon, Kümeleme Analizi

## 1. Giriş

Hemoliz, kırmızı kan hücrelerinin (alyuvar, eritrosit, RBC) zarlarının parçalanmasıyla, içlerinde bulunan hemoglobin ve diğer iç bileşenlerin hücreleri çevreleyen sıvıya (plazma veya serum, kan alma tüpünün türüne bağlı olarak) salınmasıdır. Hemoliz görsel olarak, içindeki hemoglobinden dolayı şeffaftan kırmızı renge dönen plazma sıvısından anlaşılabilir [1]. Hemoliz serum örneklerinde yaygın görülen bir durumdur ve laboratuvar test parametrelerini bozabilir. Hemoliz iki sebepten oluşabilir:

a. In-vivo hemoliz, otoimmün hemolitik anemi veya transfüzyon reaksiyonu gibi patolojik koşullar nedeniyle olabilir [1].

b. In-vitro hemoliz, uygunsuz örnek toplama, örnek işleme, ya da santrifüj nedeniyle olabilir[2].

In-vivo hemolizin tüm hemolitik örnekler içinde görülme sıklığı %3.2 oranındadır [3]. In-vitro hemoliz daha sık meydana gelir ve örneklerin taşınması veya depolanması hemoliz riskini artırır[2]. Hemoliz klinik laboratuvarlarda en yaygın görülen preanalitik hata kaynağıdır ve reddedilen test örneklerinin %40-70'e yakınının gerekçesidir [3,4]. Hemolitik örnekler, istenen tüm rutin testleri içinde %3,3 gibi bir paya sahiptir ve yetersiz numune, pıhtılı numune, doğru olmayan numune gibi ret nedenlerine kıyasla 5 kat daha fazla gözlenir [4]. Hemoliz, hücre içeriğinde bulunan potasyum (K), demir (Fe), magnezyum (Mg), aspartat amino transferaz (AST), laktat dehidrogenaz (LDH), bilirubin gibi plazma bileşenlerinde hatalı yükselme veya düşüslere yol açabilir [1].

Bu çalışmada Haziran 2011-Haziran 2012 tarihleri arasında Karadeniz Teknik Üniversitesi Tıp Fakültesi Farabi Hastanesi yatan hasta verileri kullanılarak potasyum (K) ve hemoliz indeksi (HI) arasındaki ilişki veri madenciliği tekniklerinden Kümeleme Analizi (Clustering) yöntemi kullanılarak irdelenmiştir. Bu yöntem HI üzerine yapılan bir çalışmada ilk defa kullanılmıştır ve diğer çalışmaların çoğunluğunda kan örneklerinin çalışmalar için özellikle hemolize edilmesine karşın, bu analizde gerçek hasta verileri kullanılmıştır.

## 2. Gereç ve Yöntem

### *Ön İşlemler*

Çalışmada kullanılan kan örnekleri Karadeniz Teknik Üniversitesi Tıp Fakültesi Farabi Hastanesi Haziran 2011-Haziran 2012 döneminde toplanmıştır. Belirlenen tarih aralığında kan örnekleri tekrarlanmış olan 9777 yatan hastadan elde edilen son iki potasyum (K) ve hemoliz indeksi (HI) ölçümleri çalışmada kullanılmıştır. Oluşturulan veri setinde her bir kayıt satırı, son K ve HI değerleri ve sonuç tarihini, bir önceki K ve HI değerleri ve sonuç tarihini, K ölçüm çiftinin farkını, HI ölçüm çiftinin farkını içermektedir. Elde edilen veri seti, bir ön elemeye tabi tutulmuş, kayıtlardan ilk ve son K değerleri arasındaki farklar ile ilk ve son HI değerleri arasındaki farkların her ikisinin de sıfırın üzerinde –pozitif- olduğu, ya da sıfırın altında –negatif- olduğu kayıtlar seçilmiş, diğer kayıtlar çalışma kapsamı dışında bırakılmıştır. Bu işlemin amacı, K ve HI değerleri arasındaki farkın ilişkili olduğu, yani aynı yönlü değiştiği verilerin seçilmesidir. Ön eleme adımının ardından kayıt sayısı 4394'e düşmüştür. Kalan kayıtlar alınan kan örneklerinin arasındaki süre farkı gözetilerek tekrar süzölmüş ve

kayıt sayısı 1252'ye düşmüştür.

Seçilen kayıtlar, ilk ve son K ve HI farklarının mutlak değerlerine göre kümeleme analizi uygulanarak sınıflandırılmış; daha sonra her bir sınıftaki K ve HI değerleri arasındaki korelasyon ilişkileri saptanmıştır.

### ***Kümeleme Analizi (Clustering)***

Sınıflandırma işleminden farklı olarak nesnelerin hangi sınıfa ait olduğu önceden bilinemez. Kümeleme analizi nesnelerin ortak özellikleri arasındaki mesafenin kümeler içinde minimum, kümeler dışında ise her kümenin diğerleri ile arasındaki mesafenin maksimum olmasını sağlayan bir tekniktir. Başka bir deyişle küme içindeki her bir nesne birbirine çok benzer iken, diğer küme elemanlarına ise benzememektedir. Benzerlikler veya farklılıklar probleme konu olan nesnelere tanımlayan özellikleri aracılığı ile belirlenmektedir.

Kümeleme analizi birçok çalışma alanında verideki benzer nitelikteki objeleri anlamlı kümelere ayırmak için ilk işlem olmaktadır. Küme içi benzerliklerin veya kümeler arası farkı ifade etmek için mesafe terimi kullanılmaktadır. Kümeleme analizin temeli de objeler arası mesafenin bulunmasına ve ilgili kümelerin bu mesafeye göre obje atanmasına dayanmaktadır. Mesafe ölçümü için Euclidian, Manhattan, Minkowski gibi ölçü birimleri kullanılmaktadır. Mesafeleri ölçmek için kullanılan en popüler yöntem k – Means olarak isimlendirilmektedir.

Herhangi bir nesnenin birden fazla özelliği dikkate alınarak yapılacak bir kümeleme bu özelliklerin kesikli, sürekli ya da kategorik değişken olmasına göre ya da aynı tür veri olsa bile dağılım ölçütlerinin (Ortalama, standart sapma, varyasyon katsayısı vb.) farklı olması nedeniyle kümelemeyi olumsuz etkileyecektir. Örneğin basket, güreş, yüzme vb. alanlara sporcu seçimi yapılacaksa sporcuların boy özelliğinin cm, kilo özelliğinin kilogram cinsinden verilmesi, nesnelere arası mesafeyi aynı derecede etkileyemezler. Bu nedenle kümeleme işleminden önce verilerin standardize edilmesi gerekmektedir (Eşitlik 1).

Bu çalışmada Statistica istatistiksel yazılım programı kullanılmış olup programın kümelemede kullandığı temel algoritmalar aşağıda verilmiştir (Eşitlik 2).

Veri nesnelere ait her bir özelliğin standardize edilmesi için;

$$X_j^s = \frac{X_j - X_{\text{Max}}}{\text{Max}(X_j) - \text{Min}(X_j)} \quad 1$$

Standardize edilmiş verilerin Euclidian mesafe ölçümüne göre k – Means formülü;

$$d(i, k) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{in} - X_{jn})^2} \quad 2$$

Burada  $i = (X_{i1}, X_{i2}, \dots, X_{in})$  ve  $j = (X_{j1}, X_{j2}, \dots, X_{jn})$  n boyutlu nesne başka bir deyişle n özelliği ile tanımlanan bir nesnedir. Örneğin  $X_1 = (1, 2)$  ve  $X_2 = (3, 5)$  iki boyutlu özellik içeren nesnelere olduğunu varsayarsak;

$$\text{Euclidian distance} = \sqrt{2^2 + 3^2} = 3,61$$

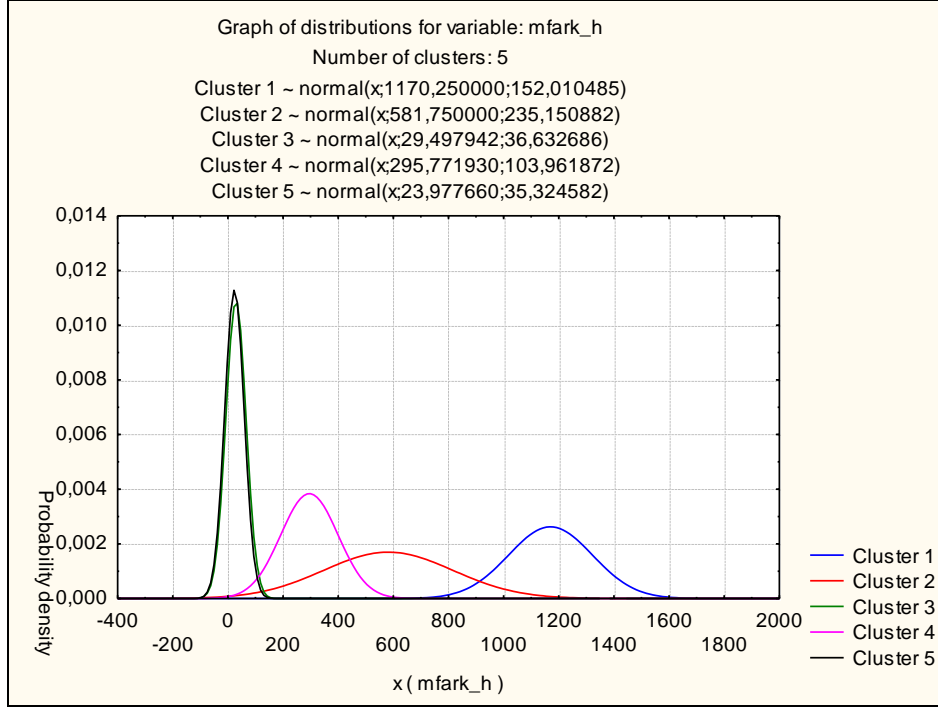
olacaktır. Dolayısıyla kümeleme işleminden önce araştırma niteliğine göre küme sayısı önceden araştırmacı tarafından belirlenmekte, küme sayısına göre her kümeye atanan ilk nesnelerin değerleri o kümenin ortalaması kabul edilerek kümeye olan mesafeler hesaplanmakta, bu hesaplama sonucunda her bir nesne en az mesafeli olduğu kümeye atanmaktadır. Yeni ortalamalar atanan bu nesnelerle tekrar hesaplanmakta ve mesafeler yeniden hesaplanmaktadır. İterasyon ortalamaların artık değişmediği ya da belirli bir hata oranında değiştiği durumda son bulmaktadır. Sonuç olarak her nesne en yakın olduğu kümede yer almış olacaktır.

### ***Korelasyon Analizi***

Basit korelasyon analizi, iki değişken arasındaki ilişkinin gücünü ve yönünü belirlemek amacı ile yapılır. Her iki değişkenin de sürekli değişken olması ve değişkenlere ilişkin verilerin normal dağılım göstermesi durumunda değişkenler arasındaki ilişki korelasyon katsayısı ile belirlenir. Fakat korelasyon değeri hiç bir şekilde neden - sonuç ilişkisi kurmaz. Korelasyon katsayısının alabileceği en küçük değer  $-1$ , en büyük değerse  $+1$  ( $-1 \leq r \leq +1$ ) olur. Katsayı, ilişkinin olmadığı durumda  $0$ , tam ve kuvvetli bir ilişki varsa  $1$ , ters yönlü ve tam bir ilişki varsa  $-1$  değerini alır [5].

### ***Modelin Kurulması ve Çalıştırılması***

Çalışmada kümeleme analizi, sürekli değişken olarak ilk ve son K ve HI farklarının mutlak değerleri tanımlanarak yapılmıştır. Yapılan tüm analizlerde Statistica istatistiksel yazılım programı kullanılmıştır. Analizde kümeler arası mesafe ölçümü için Euclidian ölçü birimi ve k-means algoritması kullanılmıştır. Kümeleme analizinde sınıf sayısı 5 olarak tanımlanmıştır. Bunun nedeni son HI değerlerine göre yapılan hemoliz sınıflandırmasıdır (HI $\leq$ 5 mg/dL için 'hemoliz yok', HI $\leq$ 30 mg/dL için 'belirsiz', HI $\leq$ 60 mg/dL için 'hafif', HI $\leq$ 200 mg/dL için 'orta', HI $>$ 200 mg/dL için 'ağır'). Analizde hata oranı 0,043588 olarak saptanmıştır. Kümeleme sonuçlarına göre gruplardaki HI fark ortalamalarına ait dağılım grafiği aşağıda verilmiştir (Grafik 1).



Grafik 1 –Gruplardaki HI fark ortalamalarına ait dağılım grafiği.

Kümele oluşturulduktan sonra her bir küme içerisinde K ve HI fark ortalamaları arasındaki korelasyon ilişkisi sorgulanmıştır. Sonuçlar aşağıdaki gibidir:

Tablo 1 – K ve HI fark değerleri ile yapılan kümeleme analizi sonuçları ve her bir kümede K ve HI fark ortalamaları arasında bulunan korelasyon katsayısı.

Kümeleler	K (Fark Ortalamaları ,p=0,00)	HI (Fark Ortalamaları,p=0,00)	n	Yüzde (%)	r
1	4,925000	1170,250	4	0,31949	0,99
2	3,750000	581,750	8	0,63898	0,80
3	1,102058	29,498	243	19,40895	0,58
4	1,445614	295,772	57	4,55272	0,91
5	0,353936	23,978	940	75,07987	0,54

### 3. Bulgular

Potasyum testi beklenen değerleri 3–6 mmol/L olarak tanımlanmaktadır [6]. Kümeleme analizi sonuçlarında da görüldüğü gibi, oluşan ilk kümede K değerleri farkı 4,925 çıkmıştır; bu sonuç grubun K değerlerinin referans değerleri aralığının çok dışında olduğunu göstermektedir. Verilerin %0,31’lik bölümünü kapsayan kümede 4 vaka

saptanmıştır. Kümenin HI değerleri fark ortalaması 1170,25 bulunmuştur. Bu kümenin istatistiklerine bakıldığında, mutlak HI farkları küme ortalaması minimum 1052, maksimum 1425 olduğu görülmektedir. Bu küme için mutlak K ve HI fark ortalamaları arasındaki korelasyon katsayısı 0,99 bulunmuştur.

Sonuçlara bakıldığında en yüksek ikinci korelasyon değeri 0,91 ile 4. kümede hesaplanmıştır. Küme kapsamında verilerin %4,55' lik kısmı olan 57 vaka kaydı vardır. K ve HI fark ortalamalarına bakıldığında, K için 1,445, HI için 295,772 olduğu görülmektedir.

3 ve 5. kümelerde yer alan 1183 (%94,47) örnek HI fark ortalamaları 30 mg/dL altında olduğu için bu kümelerdeki K değeri farklarının hemolizden kaynaklanmadığı söylenebilir. 1, 2 ve 4 kümelerde çok yüksek hemoliz olduğu gözlenmektedir. Bu kümelerin K ve HI farkı korelasyonları yüksektir.

#### 4. Tartışma ve Sonuç

Köseoğlu ve ark. [2] 16 sağlıklı gönüllüden aldıkları kan örneklerini 5 gruba ayırmış ve laboratuvar ortamında hemolize olmamış, belirsiz, hafif hemolize, orta hemolize ve ağır hemolize örnekler elde etmişlerdir. Ağır hemolize örneklerde K 5,58 mmol/L ( $p=0,002$ ) seviyesinde ölçülmüştür.

İn-vitro hemolize serum örneklerinde HI kullanılarak K yoğunluğunun tahmin edildiği bir çalışmada [8] korelasyon 0,91 olarak bulunmuştur. Bu sonuç HI farkının 295 mg/dL' den yüksek kümelerde (1,2,4) K ve HI farkları korelasyon değerleri ile uyumludur(0,99; 0,80; 0,91). 1 ve 2 nolu kümelere düşen vaka sayıları az olduğu için 4 nolu küme daha çok bilgi içermektedir.

Bu çalışmada K ve HI arasındaki ilişki kümeleme analizi yöntemi ile irdelenmiştir. K ve HI arasındaki ilişkiyi irdelleyen diğer çalışmalarda [7, 8, 9] genellikle HI'ye göre statik sınıflandırma yöntemleri kullanılmıştır. Kan örneklerinin yapay olarak hemolize edilmesi zahmetli bir çalışma sürecidir. Kümeleme yöntemi ile elde edilen sonuçlar, mevcut datadan daha kolay yararlanmayı sağlamaktadır. Ayrıca bulgular kümeleme yönteminin, potasyum ve hemolizden etkilenen diğer testler için de düzeltme faktörü geliştirilmesinde ön çalışma olarak kullanılabileceğini göstermektedir.

#### 5. Kaynakça

- [1] Lemery L. Oh, No! It's Hemolyzed! What, Why, Who, How? *Advance for Medical Laboratory Professionals*, Feb. 15, 1998: 24-25.
- [2] Koseoglu M, Hur A, Atay A, Çuhadar S. Effects of hemolysis interferences on routine biochemistry parameters. *Biochemia Medica* 2011;21(1):79-85.
- [3] Carraro P, Servidio P, Plebani M. Haemolyzed specimens: a reason for rejection or clinical challenge? *Clin Chem* 2000;46: 306-7.
- [4] Lippi G, Blanckaert N, Bonini P, Green S, Kitchen S, Palicka V, et al. Haemolysis: an overview of the leading cause of unsuitable specimens in clinical laboratories. *Clin Chem Lab Med* 2008; 46: 764-72.
- [5] Gültekin F. Korelasyon Analizi (Correlation Analysis) <http://www.fikretgultekin.com/yukseklisans/Korelasyon%20Analizi.pdf> Son Erişim: 05 Ekim 2012
- [6] T.C. Sağlık Bakanlığı Kayseri Eğitim ve Araştırma Hastanesi. Laboratuvar Test Rehberi 2011. Yayın Tarihi: Eylül 2011, Rev Tarihi: Mayıs 2012. Son Erişim: 05 Ekim 2012 [http://www.kdh.gov.tr/pdf/Lab\\_Test\\_Rehberi.pdf](http://www.kdh.gov.tr/pdf/Lab_Test_Rehberi.pdf)
- [7] Mansour MM, Azzazy HM, Kazmierczak SC.. Correction Factors for Estimating Potassium Concentrations in Samples With In Vitro Hemolysis. *Arch Pathol Lab Med*. 2009 Jun;133(6):960-6.
- [8] Shepherd J, Warner MH, Poon P, Kilpatrick ES. Use of Haemolysis Index to Estimate Potassium

- Concentration in In-Vitro Haemolysed Serum Samples. *Clin Chem Lab Med* 2006;44(7):877–879.
- [9] Jeffery J, Sharma A, Ayling RM. Detection of Haemolysis and Reporting of Potassium Results in Samples from Neonates. *Ann Clin Biochem* 2009; 46: 222–225.

## **6. Sorumlu Yazarın Adresi**

Kemal TURHAN, kturhan\_tr@hotmail.com