



İşlem karakteristik eğrisi analizi ve eğri altında kalan alanların karşılaştırılması

The analysis of receiver operating characteristic curve and comparison of the areas under the curve

Leman Tomak*, Yüksel Bek

Ondokuz Mayıs Üniversitesi, Tıp Fakültesi Biyoistatistik Anabilim Dalı, Samsun

MAKALE BİLGİLERİ

Makale Geçmişi:

Geliş 10 / 07 / 2008

Kabul 06 / 08 / 2009

* Yazışma Adresi:

Leman Tomak
Ondokuz Mayıs Üniversitesi Tıp Fakültesi,
Biyoistatistik Anabilim Dalı
Kurupelit-Samsun
e-posta: lemant@omu.edu.tr

Anahtar Kelimeler:

ROC Eğrisi
AUC
Binormal
Nonparametrik
Duyarlılık
Özgüllük

Key Words :

ROC Curve
AUC
Binormal
Nonparametric
Sensitivity
Specificity

ÖZET

Sağlık alanında bireylerin sağlam olup olmadıklarını belirlemek amacı ile kullanılan laboratuvar tekniklerine, klinik gözlemlere veya çeşitli ölçümlere bağlı olarak karara erişilen değerlendirme kurgularına tanı testleri denir. Olgulardan elde edilen ölçümlerin, eşik değerinin altında veya üstünde yer alması ile pozitif veya negatif tanı konur. İşlem Karakteristik (Receiver Operating Characteristic) Eğrisi, farklı eşik değerleri için hesaplanan, dikey eksen üzerinde doğru pozitiflik (duyarlılık) ve yatay eksen üzerinde yanlış pozitiflik (1- özgüllük) oranlarının yer aldığı bir grafikdir. ROC eğrisinin altında kalan alan (Area Under Curve), tanı testlerinin üstünlüğü için bir karşılaştırma ölçüğü olarak kullanılır. AUC ne kadar büyük ise, hastalık durumunun tahmin edilmesinde söz konusu test, o kadar iyi bir tanı testidir. AUC tahmininde genellikle iki yöntem kullanılır. Bunlar binormal ve non-parametrik yöntemlerdir. Bu derlemenin amacı, ROC eğrileri, AUC ve tahmin yöntemleri hakkında bilgi vermek ve bir hipotez testinde kullanılan iki tanı testine ait AUC'lerin karşılaştırmalarını irdelemektir.

J. Exp. Clin. Med., 2010; 27:58-65

ABSTRACT

The evaluative operations used to make a decision depending on laboratory techniques, clinical observations or various measurements to determine if individuals are healthy or not are called diagnostic tests. A positive or negative diagnosis is made by comparing the measurement with a cutoff value. A receiver operating characteristic (ROC) curve shows the characteristics of a diagnostic test by graphing the false-positive rate (1-specificity) on the horizontal axis and the true-positive rate (sensitivity) on the vertical axis for various cutoff values. The area under of a ROC curve (AUC) is a popular measure for the accuracy of a diagnostic test. A diagnostic test which has a larger AUC makes a better predict for the existence of a disease. Two methods are used to widely estimate the AUC. These include the binormal and the empirical (nonparametric) methods. The purpose of this paper is to give a reviewed information about the research of ROC curves, AUC, estimation methods and to compare AUC of two diagnostic tests using a hypothesis test.

J. Exp. Clin. Med., 2010; 27:58-65

© 2010 OMÜ Tüm Hakları Saklıdır.

1. Giriş

Tıpta alternatif tanı algoritmaları, tanı testleri ve tedavi uygulamaları ile bunların karşılaştırmalı olarak değerlendirilmesinde sorunlar vardır. Hastalıklı ve sağlıklı grup içinde hastaları sınıflandıran bir tanı yönteminin diğerinden daha iyi olup olmadığı nasıl anlaşılabilir? Bir mamogramı okuyan ve değerlendiren hekimlerin koyduğu tanıları arasında farklılık var mıdır? Radyolojideki görüntüleme yöntemlerinden biri diğerinden daha iyi midir? Hasta hikayesi bilinerek okunan bir Bilgisayarlı Tomografi

yöntemi, hasta hikayesi bilinmeden okunan bir tomografi-den daha kesin bir tanı sağlar mı? Bu örnekler çoğaltılabilir. Bu ve benzeri birçok sorunun açıklığa kavuşturulması bir karar verme sürecidir. Çözüm için kullanılan yöntemlerden biri, İşlem Karakteristik (Receiver Operating Characteristic) Eğrisi oluşturmaktır (Hanley ve McNeil, 1983; Wagner ve ark., 2002).

ROC eğrisi, istatistik karar teorisine dayanır. 1950'lerin başlarında elektronik sinyal tanımlamaları ve radar problemlerinde kullanılmaya başlanmıştır. 1960'lı

yıllarda deneysel psikolojide kullanılmıştır. 1967’de Leo Lusted adında bir radyolog tarafından tıpta kullanımı önerilmiş ve 1969 yılında da medikal görüntüleme cihazları ile ilgili karar süreçlerinde kullanılmaya başlanmıştır (Faraggi ve Reiser, 2002; Jiezhun, 2007; Metz ve ark., 1998; Taga ve ark., 2000; Van, 1998; Obuchowski, 2003).

2. Tanı Testleri

Sağlık alanında bireylerin sağlıklı olup olmadıklarını belirlemek amacı ile kullanılan laboratuvar tekniklerine, klinik gözlemlere veya hastalığa özgün cihaz ölçümlerine bağlı olarak verilen karara ve bunlarla ilgili değerlendirme bulgularına “Tanı Testleri” denir (Taga ve ark., 2000).

3. Tanısal Yeterlilik

Tanısal yeterlilik bir yönüyle, bir testin değişik sağlık durumlarını ayırt etmede, o testin vermiş olduğu bilginin kalitesinin bir ölçüsü ve sağlıklıyı hastalardan ayırt etme özelliğinin bir ölçütüdür. Testin, sağlık/ hastalık, benign/malign hastalık, tedaviye yanıt/yanıtsızlık gibi durumları ayırt edebilme ve hastalığı önceden tahmin edebilme kapasitesidir (Taga ve ark., 2000; Zweih ve Campbell, 1993; Dawson, 1993).

Diğer yönü ile tanısal yeterlilik, testin sağladığı bilginin, hastanın iyileştirilmesindeki pratik yararını ifade eder. Diğer bir ifade ile yeterliliği yüksek bir test daha düşük maliyetli, daha az yanlış sonuç veren, daha ucuz ve teknik açıdan uygulanması daha kolay olan testtir. Bu tip testlere, klinik yararı yüksek testler denir (Taga ve ark., 2000; Zweih ve Campbell, 1993; Dawson, 1993).

4. Test Performansının Değerlendirilmesi

Bir testin performansı, testin tanısal yeterliliği ya da olguları doğru olarak alt gruplara (sağlıklı/hasta vb.) ayırabilme kapasitesi ile tanımlanabilir (Taga ve ark., 2000; Zweih ve Campbell, 1993).

Herhangi bir testin performansının değerlendirilmesi için, bazı istatistiksel ölçütler hesaplanmaktadır. Bu ölçütler içinde en sık kullanılanlar, doğru pozitiflik ve yanlış pozitiflik oranları ile ifade edildikleri için çapraz tablo istatistikleri şeklinde ortaya çıkar (Taga ve ark., 2000; Zweih ve Campbell, 1993; Dawson, 1993). Bu terimler,

Tablo I. Çapraz Tablo İstatistikleri.

		Gerçek Durum		Σ
		H +	H -	
Tanı	T+	A (DP)	B (YP)	A+B
Testi	T-	C (YN)	D (DN)	C+D
Sonucu	Σ	A+C	B+D	A+B+C+D

dört gözlü tablo (Tablo I) yardımı ile aşağıdaki gibi açıklanabilir:

Duyarlılık, gerçekte hasta olanlar (H+) arasında testin pozitif sonuç verme (T+) oranıdır. Duyarlılık = $P(T+/H+) = DP/(DP+YN) = A/(A+C)$ şeklinde gösterilir (Dirican, 2001; Metz, 1978; Grove, 2006; Le, 2003; Weinstein ve ark., 2005). Özgüllük, gerçekte hasta olmayan (H-) bireylerin testlerinin negatif sonuç verme (T-) oranıdır. Özgüllük = $P(T-/H-) = DN/(DN+YP) = D/(D+B)$ şeklinde ifade edilir (Dirican, 2001; Metz, 1978; Grove, 2006; Le, 2003; Weinstein ve ark., 2005). Yanlış Pozitiflik Oranı (YPO), gerçekte hastalığa sahip olmayanlar (H-) arasında testin yanlışlıkla pozitif sonuç verme (T+) oranı olup, $YPO = P(T+/H-) = YP/(YP+DN) = B/(B+D)$ şeklindedir (Dirican, 2001; Grove, 2006). Yanlış Negatiflik Oranı (YNO), gerçekte hasta olanlar (H+) arasında testin yanlışlıkla negatif sonuç verme (T-) oranıdır ve $YNO = P(T-/H+) = YN/(YN+DP) = C/(A+C)$ şeklinde ifade edilir (Dirican, 2001; Grove, 2006). Tanı testlerinin hasta ve sağlıklıyı ayırt etmede kullanılan eşik değerine göre YPO, YNO, duyarlılık ve özgüllük değerleri farklı olacaktır (Weinstein ve ark., 2005).

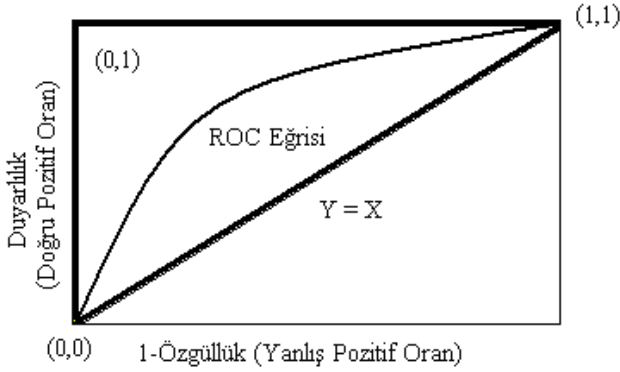
5. ROC Analizi

ROC Analizi, sadece bir duyarlılık ve özgüllük değeri kullanarak tanı koymanın getirdiği sakıncaları ortadan kaldırmak için geliştirilmiş istatistik değerlendirme yöntemidir (Metz, 1978; Metz, 2006). Bir ROC eğrisi, farklı eşik değerleri için dikey eksen üzerinde doğru pozitiflik (duyarlılık) ve yatay eksen üzerinde yanlış pozitiflik (1-özgüllük) oranlarının yer aldığı bir eğridir. ROC eğrisi üzerindeki her nokta, farklı eşik değerlerine karşılık gelen duyarlılık ve 1- özgüllük değerlerini ortaya koyar. Genelde düşük yanlış pozitiflik oranlarını veren eşik değerleri, düşük doğru pozitiflik oranına da sahiptir. Doğru pozitiflik oranı arttıkça, yanlış pozitiflik oranı da artar (Wagner, 2007; Faraggi ve Reiser, 2002; Van ve Pattynama, 1998; Zweih ve Campbell, 1993; Dirican, 2001; Grove, 2006; Weinstein ve ark., 2005; Dirican, 1991; Obuchowski, 1997; Knapp ve Miller, 1992; Obuchowski, 2005; Obuchowski, 2004).

ROC eğrisi; testin ayırt etme gücünün belirlenmesine, çeşitli testlerin etkinliklerinin kıyaslanmasına, uygun pozitiflik eşiklerinin belirlenmesine, laboratuvar sonuçlarının kalitesinin izlenmesine, uygulayıcının gelişiminin izlenmesine ve farklı uygulayıcıların tanı etkinliklerinin kıyaslanmasına olanak sağlar (Dirican, 2001).

En faydalı tanı testi, doğru pozitiflik oranı yüksek ve yanlış pozitiflik oranı düşük olan testtir. Mükemmel yakın bir tanı testi, hemen hemen dikey (0,0)’dan (0,1)’e ve sonra yatayda (1,1)’den geçen bir ROC eğrisine sahip olmalıdır (Şek.1). Kısaca sol üst köşeye en yakın geçen

ROC eğrisini veren test en kullanışlı testtir (Dirican, 2001; Weinstein ve ark., 2005; Dirican, 1991; Obuchowski, 2005; Sasse, 2002). ROC eğrisi, $Y = X$ fonksiyonuna yaklaştıkça başarısız bir test ortaya çıkar. (0,0) ile (1,1) noktalarını birleştiren köşegen çizgi referans çizgisi olarak kabul edilir (Şek.1). Bu çizgiye yakın bir ROC eğrisine sahip bir tanı testi hastalıkların taranmasında yararsız bir tanı testidir (Dirican, 2001; Weinstein ve ark., 2005; Dirican, 1991; Obuchowski, 2005; Sasse, 2002).



Şek. 1. ROC Eğrisi.

6. ROC Eğrisi Altında Kalan Alan

Bir tanı testinin tanısız yeterliliğini belirlemek için kullanılabilen pratik bir yöntem, performansın tek bir değer ile ifadesidir (Faraggi ve Reiser, 2002; Obuchowski ve McClish, 1997; Obuchowski ve ark., 2004). En yaygın kullanılan ölçüm ise, ROC eğrisinin altında kalan alandır (Area Under Curve) (Dirican, 2001; Dirican, 1991; Obuchowski, 2005; Hanley ve McNeil, 1982). AUC ne kadar büyük ise hastalığın tahmin edilmesinde test o kadar iyi olur (Şek. 2). AUC'nin olası değerleri 0.5'ten (tanı konulamaz) 1.0'e (mükemmel tanı konulabilir) kadar değişim gösterir (Grove, 2006; Hanley ve McNeil, 1982).

AUC, fiziksel bir yorumlamaya sahiptir. Hastalıklı popülasyondan seçilmiş bir kişinin kriter değerinin, hastaliksız bir popülasyondan seçilen bir kişinin kriter değerinden daha fazla olması olasılığıdır (Hanley ve McNeil, 1982). Örneğin AUC = 0.80 olduğunda, hastalıklı gruptan rasgele seçilmiş bir birey %80 olasılıkla, hastaliksız

gruptan seçilmiş bir bireye göre eşik değeri itibariyle daha yüksek test sonucu verir. Hastalık sonucunun %80'lik bir olasılıkla görüleceğini veya pozitif sonucun %80 olasılıkla hastalıkla ilişkili olacağını göstermez.

Eşik değer aynı kalsa dahi, hastalıklı ve sağlıklı popülasyonların sahip olduğu değerler birbirinden uzaklaştığında ROC eğrisindeki değişim incelenirse, hastalıklı ve sağlıklı popülasyonların kriter değerleri (grup dağılışı ortalamaları) birbirinden uzaklaştıkça ROC eğrisinin alanının arttığı görülür.

AUC'nin hesaplanması için farklı yöntemler kullanılmaktadır:

1.a. Binormal Yöntem - AUC'nin Olasılık Formülleri ile Hesaplanması

Binormal yöntem Mcclish (1989) tarafından geliştirilmiştir (McClish, 1989). Bu yöntemde, biri hastalıklı ve diğeri hastaliksız iki popülasyon olduğu ve kriter değeri değişken değerinin tüm kişiler için tanıda kullanılan bir özellik olduğu varsayılır. Hastaliksız popülasyondaki kriter değeri X ve hastalıklı gruptaki kriter değeri Y ile gösterilir ve olasılık dağılışı ifadeleri aşağıdaki şekilde gösterilebilir (Metz ve ark., 1998; Karayianni ve ark., 1996; Metz ve Pan, 1979).

Binormal modelde X (sağlıklı bireyler) ve Y (hastalıklı bireyler) değişkenlerinin farklı ortalama ve varyansa sahip ayrı iki normal dağılışı gösterdiği varsayılır (Faraggi ve Reiser, 2002; Metz ve ark., 1998; McClish, 1989; Karayianni ve ark., 1996; Metz ve ark., 1999).

$$X \approx N(\mu_X, \sigma_X^2), \quad Y \approx N(\mu_Y, \sigma_Y^2)$$

ROC eğrisi, aşağıdaki fonksiyondan yararlanılarak çizilir (Metz ve ark., 1998; Karayianni ve ark., 1996; Metz ve

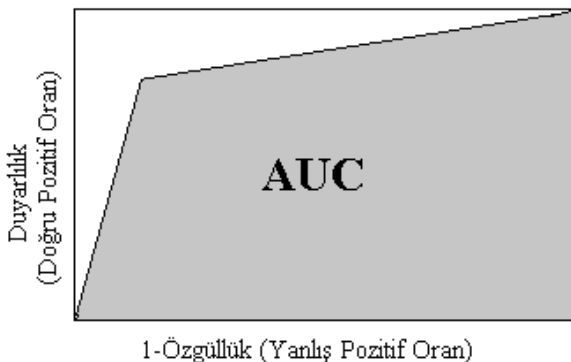
$$YPO(c) = \Phi\left(\frac{\mu_X - c}{\sigma_X}\right), \quad DPO(c) = \Phi\left(\frac{\mu_Y - c}{\sigma_Y}\right), \quad -\infty < c < \infty$$

ark., 1999).

Burada c, kriter değeridir. Doğru ve yanlış pozitiflik olasılık fonksiyonlarının birleşik fonksiyonundan AUC hesaplanabilir. Bu yöntemle çizilen ROC eğrisi düzgün (smooth) değişen bir eğridir (Metz ve ark., 1998; McClish, 1989; Karayianni ve ark., 1996; Metz ve ark., 1999).

$$\begin{aligned} A &= \int_{-\infty}^{\infty} DPO(c) YPO(c) dc \\ &= \int_{-\infty}^{\infty} \left[\Phi\left(\frac{\mu_Y - c}{\sigma_Y}\right) \Phi\left(\frac{\mu_X - c}{\sigma_X}\right) \right] dc \\ &= \Phi\left[\frac{a}{\sqrt{1+b^2}}\right] \end{aligned}$$

Kullanılan "a" ve "b" terimleri aşağıdaki gibi elde edilir (Metz ve ark., 1998; McClish, 1989; Karayianni ve ark., 1996; Metz ve ark., 1999).



Şek. 2. ROC Eğrisi Altında Kalan Alan (AUC).

$$a = \frac{\mu_y - \mu_x}{\sigma_y} = \frac{\Delta}{\sigma_y}, \quad b = \frac{\sigma_x}{\sigma_y}, \quad \Delta = \mu_y - \mu_x$$

Alanın varyansı aşağıdaki diferansiyel yöntemler kullanılarak elde edilir (McClish, 1989);

$$v(\hat{A}) = \left(\frac{\partial A}{\partial \Delta}\right)^2 v(\hat{\Delta}) + \left(\frac{\partial A}{\partial \sigma_x^2}\right)^2 v(S_x^2) + \left(\frac{\partial A}{\partial \sigma_y^2}\right)^2 v(S_y^2)$$

$$\frac{\partial A}{\partial \Delta} = \frac{E}{\sqrt{2\pi(1+b^2)}\sigma_y^2} [\Phi(\tilde{c}_1) - \Phi(\tilde{c}_0)]$$

$$\frac{\partial A}{\partial \sigma_x^2} = \frac{E}{4\pi(1+b^2)\sigma_x\sigma_y} \left[e^{-k_0} - e^{-k_1} \right] - \frac{abE}{2\sigma_x\sigma_y\sqrt{2\pi(1+b^2)}^{3/2}} [\Phi(\tilde{c}_1) - \Phi(\tilde{c}_0)]$$

$$E = \exp\left(-\frac{a^2}{2(1+b^2)}\right)$$

$$\frac{\partial A}{\partial \sigma_y^2} = -\frac{a}{2\sigma_y} \left(\frac{\partial A}{\partial \Delta}\right) - b^2 \left(\frac{\partial A}{\partial \sigma_x^2}\right)$$

$$\tilde{c}_1 = \left[\Phi^{-1}(FP_1) + \frac{ab}{(1+b^2)} \right] \sqrt{(1+b^2)}$$

$$k_i = \frac{\tilde{c}_i^2}{2}$$

$$V(\Delta) = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$$

$$V(S_x^2) = \frac{2\sigma_x^4}{n_x - 1}$$

$$V(S_y^2) = \frac{2\sigma_y^4}{n_y - 1}$$

1.b. Binormal Yöntem - AUC'nin Çift Yönlü Olasılıklı Grafik Kağıdı ile Hesaplanması

Pratik olarak kullanılabilir bir yöntemdir. Bu yöntemde, her iki eksen de olasılık işaretli grafik kağıdı üzerinde binormal (X ve Y dağılımları için) grafik üzerindeki regresyon hattının eğim ve kesim noktalarından yararlanılarak alan hesabı yapılabilir (Hanley ve McNeil, 1983; Swets, 1979):

$$A = \phi \left[\frac{a}{\sqrt{1+b^2}} \right]$$

Formülde gösterilen "a", ROC eğrisinin kesim noktasını ve "b", ROC eğrisinin eğimini ifade eder (Hanley ve McNeil, 1983; Swets, 1979).

2. Nonparametrik Yöntem ile AUC Hesabı

Nonparametrik yöntem, AUC hesaplanmasında kullanılan bir diğer yöntemdir (Hanley ve McNeil, 1983; Hanley ve McNeil, 1982; DeLong ve ark., 1988). Özellikle binormal yöntemin yaptığı güçlü normallik varsayımlarını karşılamadığı için bu yöntem popüler olmuştur (DeLong ve ark., 1988).

ROC eğrisi altında kalan alan aslında bir olasılık ölçüsüdür. Bu alanın olasılıkla ifade edilen değeri "θ" simgesi ile tanımlanabilir. Rasgele seçilen sağlam ve hasta denek veya normal ve anormal görüntü eşleştirildiğinde, bu iki denek veya görüntü değerlerindeki benzeşim bunların doğru sınıflandırılmalarına yardımcı olur. Yani, ROC eğrisi altındaki gerçek alan q'ya eşittir ve q = P(XA > XN) şeklinde gösterilir (Hanley ve McNeil, 1982). Rasgele seçilen normal ve anormal değerlerin doğru olarak sıralanma olasılığı Wilcoxon istatistiği (W = θ olasılığı = eğri altında kalan alan) ile ölçülür (Hanley ve McNeil, 1982).

A grubunun örnek büyüklüğü nA ve N grubunun örnek büyüklüğü nN ile gösterildiğinde, bu iki grubun değerleri arasındaki olası tüm karşılaştırmaların sayısı nA * nN'den oluşmaktadır. Kurala göre her karşılaştırmaların skorlanması ve bu skorlar kullanılarak alanın hesaplanması aşağıdaki gibidir: Gruplardan birinin puanları sıraya dizildikten sonra, diğer grubun puanlarının, bu grubun puanlarına göre küçük, eşit veya büyük oluşuna göre S'nin aldığı değer 1, 0.5 ve 0 olur. Eğer X özelliği iyi bir ayırım yeteneğine sahipse, bu olasılık 1'e daha yakın olacaktır, değilse 0.5'e yakın olacaktır (Dirican, 2001; Dirican, 1991; Hanley ve McNeil, 1982; DeLong ve ark., 1988). Bunun sonrasında da W (θ olasılığı) hesaplanabilir (Hanley ve McNeil, 1982; DeLong ve ark., 1988).

$$\text{eger } x_A > x_N \rightarrow 1$$

$$S(x_A, x_N) = \begin{cases} \text{eger } x_A > x_N \rightarrow 1 \\ \text{eger } x_A = x_N \rightarrow 1/2 \\ \text{eger } x_A < x_N \rightarrow 0 \end{cases} \quad W = \hat{\theta} = \frac{1}{n_A \cdot n_N} \sum_{i=1}^{n_A} \sum_{j=1}^{n_N} S(x_{A_i}, x_{N_j})$$

$$\text{eger } x_A < x_N \rightarrow 0$$

θ'ya ait standart hata [SH(θ)] hesabı yapılırken; Q1 ve Q2 değerleri kullanılır.

SH hesabı için gerekli olan Q1 ve Q2'nin elde edilmesi aşağıdaki gibi ifade edilir (Obuchowski, 2003; Hanley ve McNeil, 1982):

$$Q_1 = \theta \div (2 - \theta)$$

$$Q_2 = 2\theta^2 \div (1 + \theta)$$

Standart hata ise:

$$SH = \sqrt{\frac{\theta(1-\theta) + (n_A - 1)(Q_1 - \theta^2) + (n_N - 1)(Q_2 - \theta^2)}{n_A n_N}}$$

Tek Bir ROC Eğrisinin Değerlendirilmesi

Bir tanı testinin kullanılabilirliğine ait istatistik test onu $H_0: A=0,5$ değeri ile karşılaştırmaktır. Tahmini AUC'nin testinde normal dağılım yaklaşımının kullanılabilmesi için örneğin yeteri kadar büyük olduğu varsayılır ve böylece Z istatistiği ile bir test yapılabilir (Hanley ve McNeil, 1993; McClish, 1989).

İstatistiksel test: $Z = \frac{A-0,5}{\sqrt{V(A)}}$ şeklindedir.

A, AUC'nin tahmini ve $V(A)$, A'nın varyansının tahminidir (Hanley ve McNeil, 1993).

İki ROC Eğrisinin AUC'lerinin Karşılaştırılması

İki yöntemin tanısal yeterliliğini karşılaştırmak için eğri altında kalan alanlar karşılaştırılabilir. İki ROC eğrisine ait alanların benzer olması eğrilerin farklı olmadığı anlamına gelmemektedir. Bu nedenle istatistiki olarak bu alanları karşılaştırmak gerekir (Hanley ve McNeil, 1993; McClish, 1989; DeLong ve ark., 1988).

1. İki Binormal ROC Eğrisinin AUC'lerinin Karşılaştırılması

Binormal varsayım uygulandığında, iki ROC eğrisinin altındaki alanlara ait hipotezlerin eşitliği aşağıdaki formül kullanılarak test edilebilir (McClish, 1989).

$$Z = \frac{A_1 - A_2}{\sqrt{V(A_1 - A_2)}}$$

Eğer Z değeri, kritik bir seviyenin üzerinde ise, iki alan arasında fark olduğu kabul edilir. Genellikle kritik seviye $\alpha = 0,05$ düzeyinde 1,96 olarak kabul edilir ve sonrasında hesaplanan değer kritik değerden büyükse iki eğrinin benzer olduğu şeklindeki hipotez reddedilir (McClish, 1989).

Bağımsız Örnekler: Bağımsız örnekler tasarımı

$$V(A_1 - A_2) = V(A_1) + V(A_2)$$

kullanıldığında, farklı AUC'lerin varyansları, kovaryans sıfır olduğu için varyansların toplamıdır. Bu,

$$V(A_1 - A_2) = V(A_1) + V(A_2) - 2Cov(A_1 A_2)$$

şeklinde hesaplanır (McClish, 1989).

Bağımlı Örnekler: AUC'ye ait farkın varyansı, şeklindedir (McClish, 1989).

Veriler bağımlı olduğunda kovaryans teriminin kullanılması gerekir ve aşağıdaki gibi hesaplanır:

$$Cov(A_1, A_2) = \left(\frac{\partial A_1}{\partial \Delta_1} \right) \left(\frac{\partial A_2}{\partial \Delta_2} \right) Cov(\hat{\Delta}_1, \hat{\Delta}_2) + \left(\frac{\partial A_1}{\partial \sigma_{x_1}^2} \right) \left(\frac{\partial A_2}{\partial \sigma_{x_2}^2} \right) Cov(s_{x_1}^2, s_{x_2}^2) + \left(\frac{\partial A_1}{\partial \sigma_{y_1}^2} \right) \left(\frac{\partial A_2}{\partial \sigma_{y_2}^2} \right) Cov(s_{y_1}^2, s_{y_2}^2)$$

$$Cov(\hat{\Delta}_1, \hat{\Delta}_2) = \frac{\rho_x \sigma_{x_1} \sigma_{x_2}}{n_x} + \frac{\rho_y \sigma_{y_1} \sigma_{y_2}}{n_y}$$

$$Cov(s_{x_1}^2, s_{x_2}^2) = \frac{2\rho_x \sigma_{x_1}^2 \sigma_{x_2}^2}{n_x - 1}, \quad Cov(s_{y_1}^2, s_{y_2}^2) = \frac{2\rho_y \sigma_{y_1}^2 \sigma_{y_2}^2}{n_y - 1}$$

2. İki Nonparametrik ROC Eğrisinin AUC'lerinin Karşılaştırılması

Varyans ve kovaryans formülleri, tasarımın bağımlı ya da bağımsız örnek olup olmamasına bağlı olarak kullanılır. İki AUC'nin karşılaştırılması için kullanılan Z testi aşağı-

$$Z = \frac{A_1 - A_2}{\sqrt{V(A_1 - A_2)}}$$

$$V(A_1 - A_2) = V(A_1) + V(A_2) - 2Cov(A_1 A_2)$$

da verilmiştir (Hanley ve McNeil, 1983; DeLong ve ark., 1988; Zhou ve ark., 2002).

$$V(A_k) = \frac{S_{T_{k1}}}{n_{k1}} + \frac{S_{T_{k0}}}{n_{k0}}$$

$$S_{T_{ki}} = \frac{1}{n_{ki} - 1} \sum_{j=1}^{n_{ki}} [V(T_{kij}) - A_k]^2, \quad k=1,2 \quad i=0,1$$

$$V(T_{k1i}) = \frac{1}{n_{k0} - 1} \sum_{j=1}^{n_{k0}} \psi(T_{k1i} - T_{k0j}), \quad k=1,2$$

$$V(T_{k0j}) = \frac{1}{n_{k1} - 1} \sum_{i=1}^{n_{k1}} \psi(T_{k1i} - T_{k0j}), \quad k=1,2$$

$$A_k = \frac{\sum_{i=1}^{n_{k1}} V(T_{k1i})}{n_{k1}} = \frac{\sum_{j=1}^{n_{k0}} V(T_{k0j})}{n_{k0}}, \quad k=1,2$$

$$\psi(X, Y) = \begin{cases} 0 & \rightarrow \text{eger } Y > X \\ 1/2 & \rightarrow \text{eger } Y = X \\ 1 & \rightarrow \text{eger } Y < X \end{cases}$$

Bağımsız Örnekler: Her deneye iki tanı testinden sadece biri uygulandığı için kovaryans sıfırdır ve iki varyans toplanır (DeLong ve ark., 1988; Zhou ve ark., 2002):

Bağımlı Örnekler: Her deneye iki tanı testinin ikisi de uygulandığından A_1 ve A_2 'nin farkının varyansı hesaplanırken, grupların bağımlı olması nedeniyle kovaryans teri-

$$\text{Cov}(A_1, A_2) = \frac{S_{T_{11}T_{21}}}{n_1} + \frac{S_{T_{10}T_{20}}}{n_0}$$

$$S_{T_{11}T_{21}} = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} [V(T_{11j}) - A_1][V(T_{21j}) - A_2]$$

$$S_{T_{10}T_{20}} = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} [V(T_{10j}) - A_1][V(T_{20j}) - A_2]$$

mi de dikkate alınacağından, varyansların formülü aşağıdaki gibi elde edilir (DeLong ve ark., 1988; Zhou ve ark., 2002):

Normal dağılışa yaklaşımı sağlamak amacıyla transformasyon yapılabilir. Bu dönüşümde değişkenlerin ortalama

$$\theta(A) = \ln \left(\frac{FP_2 - FP_1 + A}{FP_2 - FP_1 - A} \right)$$

$$A = (FP_2 - FP_1) \frac{e^\theta - 1}{e^\theta + 1}$$

$$V(\theta) = \left(\frac{2(FP_2 - FP_1)}{(FP_2 - FP_1)^2 - A^2} \right)^2 V(A)$$

$$\text{Cov}(\theta_1, \theta_2) = \frac{4(FP_2 - FP_1)^2}{\left[(FP_2 - FP_1)^2 - A_1^2 \right] \left[(FP_2 - FP_1)^2 - A_2^2 \right]} \text{Cov}(A_1, A_2)$$

ve varyansları alanlar (A1, A2) ve yanlış pozitiflik oranları (FP1, FP2) kullanılarak, aşağıdaki şekilde bulunur (McClish, 1989):

Optimal Kriter Değerin Bulunuşu

Optimal kriter değer, ortalama maliyeti küçülten değerdir. Bu yaklaşım, Metz (1978) ile Zhou ve ark., (2002) tarafından geliştirilmiştir.

Bu yaklaşım bir tanı testinin dört olası sonucunun (DP: Doğru pozitif, DN: Doğru negatif, YP: Yanlış pozitif ve YN: Yanlış negatif) maliyet analizini kapsar. Bu sonuçların her birinin maliyeti belirlenmiş olmalıdır. Bu küçük bir işlem değildir. Aslında tüm çalışma alanı bu maliyetleri belirlemek maksadıyla ortaya çıkmıştır.

$$C = C_O + C_{DP} P(DP) + C_{DN} P(DN) + C_{YP} P(YP) + C_{YN} P(YN)$$

İlk önce bu maliyetler bulunur, sonra bir test geliştirilmesine ait ortalama tüm maliyet (C) hesaplanır (Metz, 1978; Zhou ve ark., 2002) :

Burada C_O geliştirilen testin sabit maliyetidir, C_{DP} doğru pozitifle ilişkili maliyettir ve $P(DP)$ popülasyondaki doğru pozitiflerin oranıdır. $P(DP)$ 'nin hesaplanması,

$P(DP) = \text{Duyarlılık} [P(\text{Durum}=\text{Doğru})]$ şeklindedir.

$$m = \frac{P(\text{Durum} = \text{Yanlış})}{P(\text{Durum} = \text{Doğru})} \left(\frac{C_{YP} - C_{DN}}{C_{YN} - C_{DP}} \right) \text{ veya } m = \frac{P(H-)}{P(H+)} \left(\frac{C_{YP} - C_{DN}}{C_{YN} - C_{DP}} \right)$$

Metz, ROC eğrisi boyunca [duyarlılık – m (1-özellik)] değerini maksimize eden ortalama maliyetin minimum olduğunu göstermiştir. Burada m değeri,

Eşitlikteki maliyet, hastalısız bir kişinin testinin net maliyetinin hastalıklı bir kişinin testinin net maliyetine oranı şeklinde yer alır. Testlerin yukarıdaki bağlamda kullanımında, optimum eşik değere ait kesim noktası, eşik değer değişkeninin her değeri için C'yi gösteren bir raporun elde edilmesi ile bulunabilir.

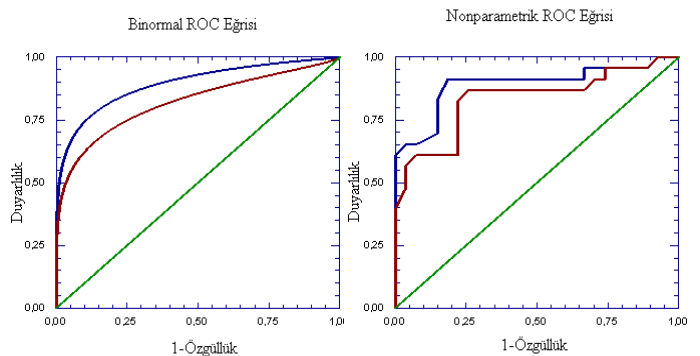
Eşik değer değiştiğinde negatif ve pozitif tanı konan hasta sayıları değişecektir. Bu değişimde kazanç ve kayıp ne olur? Hastalara sağlıklı demek mi, yoksa sağlıklılara hasta demek mi daha fazla kayba neden olur? Güvenli bir eşik değeri nasıl belirlenir? Yanlış pozitif bulmanın neticesi ne kadar kötü veya yanlış pozitif bulmanın neticesi sağlık sistemini ve hastayı ne kadar etkiler? Bunların cevaplarının bilinmesi, eşik değerin belirlenmesinde etken olmaktadır. ROC eğrisi, araştırmacı-test ikilisinin performansını birlikte yansıtır. Eğer araştırmacının hastanın göreceği zararı veya sağlık sisteminin uğrayacağı zararı kestirme bilgisi zayıfsa, ROC eğrisi ile tanı koyma performansı düşük olacak, aynı şekilde testin ayırt etme gücü zayıfsa yine performans düşecektir. Çünkü eşik değer seçimi subjektif bir olaydır.

Araştırmacı-test ikilisinin performansını değerlendirmenin yolu, maliyet analizi yapmaktır.

Tanı testi performansı değerlendirirken kişilere hasta ve sağlıklı demenin bir bedeli vardır. Bu bedel, maliyet ve komplikasyon riski şekillerinde olabilir. Yanlış karar verme maliyetine karşı doğru karar vermenin sağladığı kazanç nasıl dengelenebilir, sağlanan bilginin bedeli, ödenen maliyete değer midir ve tanı testlerinin kombinasyonu nasıl elde edilebilir sorularının cevaplanması için maliyet-fayda analizi yapılır.

Tanı Testlerinin ROC Analizi ile Karşılaştırılması

Hipertansiyon saptanması için kullanılan iki tanı testinin etkinlikleri, binormal ve nonparametrik ROC eğrisi kullanılarak karşılaştırıldı (veriler simülasyon ile elde edildi). Binormal ve nonparametrik yöntemlere ait ROC eğrileri, Şekil 3'te gösterildiği gibidir (NCSS, 2007):



Şek. 3. Binormal ve Nonparametrik Yöntemlere Ait ROC Eğrileri.

Her iki teste ait sonuçların $AUC > 0.5$ hipotezine göre Number Cruncher Statistical Systems istatistik paket programı

Tablo II. Yöntemlerin AUC>0.5 Hipotezine Göre Değerlendirilmesi.

Binormal Yöntem							
Kriter	AUC	SH	Z değeri	Tek yönlü p	İki yönlü p	Prevalans	Sayı
Tanı 1	0.893	0.046	3.93	<0.001	<0.001	0.460	50
Tanı 2	0.820	0.061	3.28	<0.001	<0.01	0.460	50
Nonparametrik Yöntem							
Kriter	AUC	SH	Z değeri	Tek yönlü p	İki yönlü p	Prevalans	Sayı
Tanı 1	0.892	0.051	7.62	<0.001	<0.001	0.460	50
Tanı 2	0.833	0.062	5.39	<0.001	<0.001	0.460	50

ile değerlendirilmesi Tablo II'de verildi (NCSS, 2007).

Bu iki tanı testine ait ROC eğrileri binormal ve nonparametrik yöntemlerle oluşturularak AUC'leri saptandı. AUC değerlerinin, her iki tanı testi için 0,5'ten daha büyük oldukları istatistiksel olarak (binormal ve nonparametrik yöntemler ile) gösterildi ($p<0,05$). Buna göre her iki tanı testi yöntemi de hipertansiyon tanısında kullanılabilir.

Tanı testlerinin, binormal ve nonparametrik yön-

Tablo III. Yöntemlerin (AUC1 - AUC2) = 0 Hipotezine Göre Değerlendirilmesi.

Binormal Yöntem							
Kriter	AUC1	AUC2	Fark	Fark SH	% Fark	Z değeri	p
Tanı 1,2	0.893	0.820	0.074	0.031	-8.12	2.48	<0.05
Tanı 2,1	0.820	0.893	-0.073	0.036	8.84	-2.48	<0.05
Nonparametrik Yöntem							
Kriter	AUC1	AUC2	Fark	Fark SH	% Fark	Z değeri	p
Tanı 1,2	0.892	0.833	0.06	0.036	-6.68	1.67	<0.05
Tanı 2,1	0.833	0.892	-0.06	0.036	7.16	-1.67	<0.05

temlere göre elde edilen alanlarına dair, (AUC1 - AUC2) = 0 hipotezi için NCSS' paket programı ile karşılaştırılması Tablo III'te verildi (NCSS, 2007).

Bu iki tanı testine ait sonuçların binormal yöntemle karşılaştırılmasında, iki testin tanı koyabilme açısından birbirinden istatistiksel olarak farklı olduğu ($p<0,05$) saptandı. Ancak, nonparametrik yöntem ile yapılan karşılaştırmada bu iki tanı testi arasında istatistiksel olarak önemli bir fark olmadığı görüldü ($p>0,05$).

Her iki tanı testinin farklı eşik değerleri için bi-

KAYNAKLAR

- Dawson, N.V. 1993. Physician judgment in clinical settings: methodological influences and cognitive performance. Clin. Chem. 39,1468-1480.
- Dirican, A. 1991. ROC eğrisi çözümlenmesi ile tanı testlerinin değerlendirilmesi ve bilgisayar uygulaması. Doktora tezi.
- Dirican, A. 2001. Tanı testi performansının değerlendirilmesi ve kıyaslanması. Cerrahpaşa Tıp Dergisi. 32, 25-30.
- DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 44, 837-845.
- Faraggi, D., Reiser, B. 2002. Estimation of the area under the ROC curve. Stat Med. 21, 3093-3106.
- Grove, W.M. 2006. Mathematical aspects of diagnosis. United States of America, 50-75.
- Hanley, J.A., McNeil, B.J. 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 148, 839-843.
- Hanley, J.A., McNeil, B.J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 143, 29-36.
- Jiezhun, G. and Subhashis Ghosal Department of Statistics. 2007. Bayesian ROC curve estimation under binormality using a partial likelihood based on ranks. North Carolina State University, Raleigh, North Carolina.
- Karayianni, T., Tretiak, O.J., Herrmann, N. 1996. ROC analysis: Comparison between the binormal and the Neyman-Pearson Model. Signals, Systems and Computers, 1996 Conference Record of the Thirtieth Asilomar Conference. 3-6 Nov 1996.
- Knapp, R.G., Miller, M.C. 1992. Clinical Epidemiology and Biostatistics. United States Of America, Williams&Wilkins Press,
- Le, C.T. 2003. Introductory Biostatistics. United States of America, Wiley & Sons Pres, 336-337.

Tablo IV. Maliyet-Fayda Analizi Sonuçları.

Eşik değeri	Maliyet-fayda değerleri (C değerleri)			
	Binormal yöntem		Nonparametrik yöntem	
	I.Tanı	II.Tanı	I.Tanı	II.Tanı
110	-7.910	-7.999	-6.379	-8
120	-6.632	-7.904	-2.420	-8
130	-2.608	-6.439	-0.014	-6.710
140	0.162	-2.193	0.348	-2.464
150	0.332	0.175	0.130	0.145
170	0.11	0.212	0	0.174
180	0.019	0.059	0	0.087

normal ve nonparametrik yöntemlere göre elde edilen maliyet-fayda analizi Tablo IV'te verildi.

Söz konusu maliyet fayda oranı, hastalık yokluğundaki net maliyetin hastalık varlığındaki net maliyete oranıdır. Hesaplanan maliyet değerinin en yüksek ya da en düşük şekilde gösterilmesi için belirli bir eşik değer seçilir. Sonucun gerçek hayata uygun olabilmesi için seçilecek maliyet fayda oranının gerçekçi olması gerekir. Bu nedenle söz konusu hastalıkla ilgili maliyet fayda oranının gerçek veriler üzerinden analizinin yapılması gerekmektedir.

Binormal yöntemde eğriler düzleştirildiği için ROC eğrisi altında kalan alan değişmektedir. Nonparametrik yöntemde ise eğri düzleştirilmeden gerçek alan hesaplanmaktadır. Bu nedenle ve varsayımlardan uzak yaklaşımlara sahip olduğu için nonparametrik yöntem, binormal yöntemle tercih edilir.

Sonuç olarak; ROC eğrisi hesaplama teknikleri basittir, görsel açıdan izlenmesi kolaydır, tüm sınırlara ilişkin testin tanısallık yeterliliğini ve doğruluğunu göstermektedir. Tek bir tanısallık karar eşliğinin seçilmesini gerektirmez ve prevalanstan bağımsızdır. Testler arasında direkt görsel kıyaslamaya olanak verebilecek özelliktedir. Histogramlarda olduğu gibi, sonuçların gruplandırılarak değerlendirilmesini gerektirmez ve duyarlılık ve özgüllük değerleri grafikten doğrudan elde edilebilir (Taga ve ark., 2000).

- Metz, C.E., Herman, B.A., Shen, J.H. 1998. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Stat Med.* 17,1033-1053.
- Metz, C.E. 1978. Basic principles of ROC analysis. *Semin Nucl. Med.* 8, 283-298.
- McClish, D.K. 1989. Analyzing a portion of the ROC curve. *Med. Decis Making.* 9,190-195.
- Metz, C.E. 1999. Pan X. "Proper" Binormal ROC Curves: Theory and Maximum- Likelihood Estimation. *J. Math. Psychology.* 43, 1-33.
- Metz, C.E. 2006. Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. *J. Am. Coll Radiol.* 3, 413-422.
- NCSS İstatistik Paket Programı. NCSS 2007 Manuals, 329 North 1000 East, Kaysville, Utah, 84037, USA.
- Obuchowski, N.A. 2003. Receiver operating characteristic curves and their use in radiology. *Radiology.* 229, 3-8.
- Obuchowski, N.A., McClish, D. 1997. Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. *Stat Med.* 16, 1529-1542.
- Obuchowski, N.A. 2005. ROC analysis. *AJR Am. J. Roentgenol.* 184, 364-372.
- Obuchowski, N.A., Lieber, M.L., Wians, F.H. 2004. ROC curves in clinical chemistry: uses, misuses, and possible solutions. *Clin. Chem.* 50,1118-1125.
- Sasse, E.A. 2002. Objective evaluation of data in screening for disease. *Clin. Chim. Acta.* 315, 17-30.
- Swets, J.A. 1979. ROC analysis applied to the evaluation of medical imaging techniques. *Invest Radiol.* 14, 109-121.
- Taga, Y., Aslan, D., Güner, G., 2000. Tıbbi laboratuvarlarda standardizasyon ve kalite yönetimi. Ankara, Türk Biyokimya Derneği Yayınları. 106-123.
- Van Erkel, A.R., Pattynama, P.M.T. 1998. Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. *Eur. J. Radiol.* 27, 88-94.
- Zweih, M.H., Campbell, G. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.* 39, 561-577.
- Zhou, X., Obuchowski, N., McClish, D. 2002. *Statistical Methods in Diagnostic Medicine.* John Wiley & Sons, Inc. New York.
- Wagner, R.F., Metz, C.E., Campbell, G. 2007. Assessment of medical imaging systems and computer aids: a tutorial review. *Acad. Radiol.* 14, 723-748.
- Weinstein, S., Obuchowski, N.A., Lieber, M.L. 2005. Clinical evaluation of diagnostic tests. *AJR Am. J. Roentgenol.* 184, 14-19.