

ROC Analizi II: ROC Eğrisinin Bilgisayar Destekli İstatistik Programları Aracılığıyla Çizilmesi

Sait Uluç

Hacettepe Üniversitesi, Psikoloji Bölümü
psysait@hacettepe.edu.tr

Bu yazının amacı, "ROC Analizi I" başlıklı yazıda temel mantığı ve elle hesaplama yöntemleri aktarılan ROC analiz yönteminin paket programlarla yapılan uygulamalarına yönelik kısa bir bilgilendirme sunmaktır. ROC eğrisinin, Microsoft Excell'de dahil olmak üzere, birçok bilgisayar paket programıyla hesaplanıp, çizilebilmesi mümkündür. Bu yazı kapsamında en yaygın kullanıma sahip olan SPSS ve MedCalc paket programları ele alınmıştır.

MedCalc

Veri girişi:

MedCalc Paket programıyla ROC analizi yapabilmek için en az iki tür veriye ihtiyaç vardır: (1) üzerinde çalıştığınız parametreye ilişkin olarak ilgili testten elde edilen sürekli puanlar (örn.; Beck Açılan

diyalog kutusunda doldurulması gereken alanlar aşağıda kısaca tanımlanmıştır. Depresyon Envanteri (BDE) puanları gibi) ve (2) katılımcılarınızı ilişkisiz iki gruba ayıran (hasta ve normal gibi) bağımsız bir tanı sınıflaması (örn.; BDE puanlarından bağımsız olarak psikiyatrik değerlendirmeye dayalı olarak yapılan bir tanı sınıflaması gibi). Veri sayfasında biri tanıya ilişkin sınıflamaları, diğeri ise test sonucu elde edilen puanları yazmak için iki sütun açılır. Bütün katılımcılara tanı için sabit bir kod atanır: Tanı alanlar için 1 ve tanı almayanlar için 0. Testten elde edilen puanlar ise uygun bir başlık altında diğer sütuna yazılır.

Komutlar:

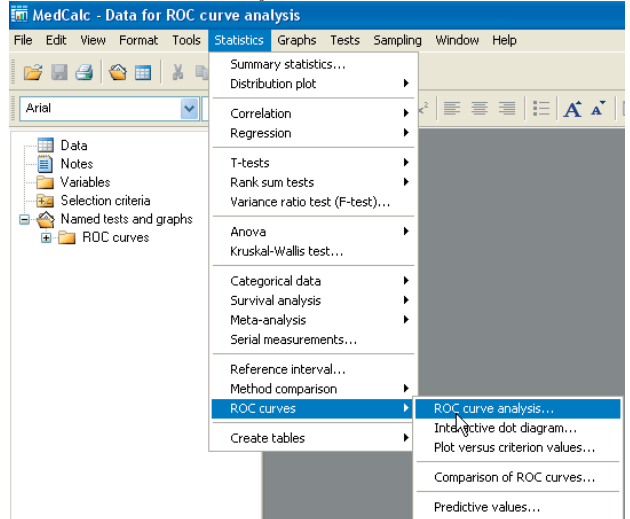
Aşağıda gösterilen komut sistemi takip edilerek diyalog kutusu açılır.

Data for ROC curve analysis		
	A	B
	TANI	TEST1
1	1	112,7
2	1	104,0
3	0	88,9
4	1	126,7
5	1	123,3
6	1	120,5
7	1	130,3
8	1	129,6
9	0	97,9
10	0	94,9
11	1	140,2
12	1	119,7
13	0	98,6
14	0	77,3
15	1	139,9
16	0	97,9
17	1	134,2
18	1	137,5
19	1	131,2

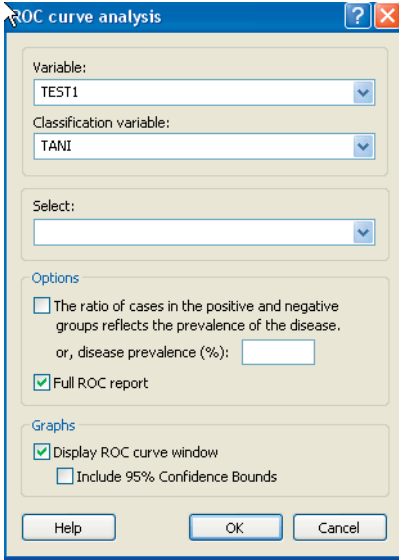
Statistics

..... ROC curves

..... ROC curve analysis



Açılan diyalog kutusunda doldurulması gereken alanlar aşağıda kısaca tanımlanmıştır.



Variable: üzerinde çalıştığımız değişkeni sağdaki kutucuğu tıklayarak buraya girin.

Classification Variable: Katagorik değişkeninizi sağdaki kutucuğu tıklayarak buraya girin (0 = tanı yok, 1 = tanı var). Eğer tanı değişkenini önceden 0 ve 1 olarak girmediyseniz, program uygun şekilde çalışmayacaktır. Bu durumda IF fonksiyonu kullanılarak bütün değerler 0 ve 1'e dönüştürülebilir (örn.; IF(RESULT="pos",1,0).

Select: (İsteğe bağlı) örneklem içindeki belli bir alt grubu analiz etmek istiyorsanız, seçim kriterini bu bölüme giriniz (örn.; YAŞ>21, CİNSİYET="Erkek").

Options:

- **Disease prevalence:** Eğer tanı alan ve almayan grupların örneklem içindeki dağılımı hastalığın gerçekteki görülme sıklığını yansıtmıyorsa, bu durum diyalog kutusunda belirtilmelidir. Hastalığın görülme sıklığının yüzdesi "Disease prevalence" kutucuğuna girilir (örn.; görülme sıklığı % 5 olan bir hastalık için kutucuğa 5 yazılır). Hastalığın görülme sıklığı ile ilgi-

li bilginiz yoksa ya da yaptığınız analizle ilişkili olmadığı düşünülüyorsa, bu bölüm boş bırakılabilir. Bu durumda program yordayıcı değerleri hesaplamaz.

- **Full ROC report:** Eğer olası tüm eşik değerler için duyarlılık ve özgüllük değerlerinin hesaplanmasını istiyorsanız, bu kutucuğun işaretlenmesi gerekir. Bu seçenek işaretlenmediğinde, program sadece ROC eğrisinin çiziminde kullanılacak önemli değerleri ve kesim noktası olarak belirlenen eşik değerdeki en yüksek duyarlılık ve özgüllük değerlerini verir.

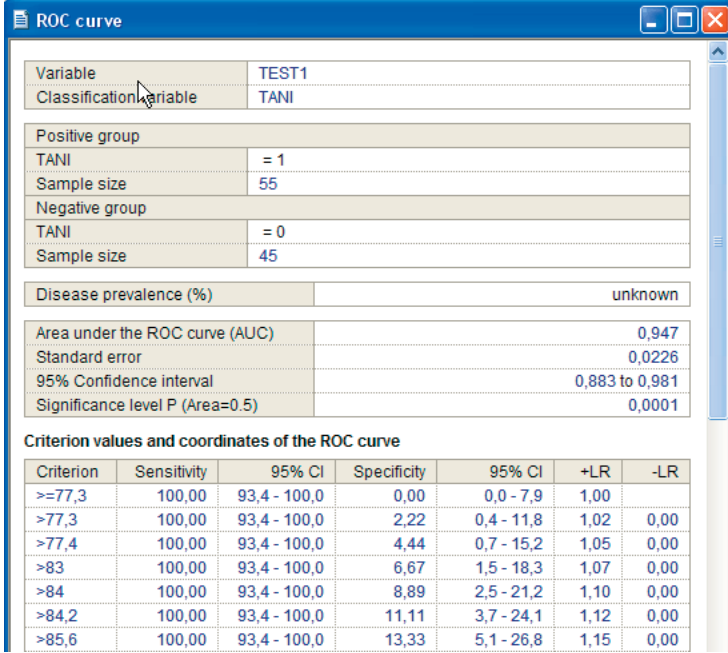
Graphs:

- **"Display ROC curve window"** seçeneği işaretlendiğinde program elde edilen ROC eğrisini ayrı bir pencerede verir.

Sonuçlar:

Sonuç sayfasındaki ilk bölüm tanı alma ölçütüne göre örneklemdaki gözlem sayılarını verir. Örnek sayfada tanı alan 55 ve tanı almayan 45 katılımcı olduğu görülmektedir. İdeal bir analiz için 50'si tanı alan ve 50'si tanı almayan gruptan olmak üzere en az 100 katılımcının bulunması gerektiği genel olarak kabul görmektedir. ROC Eğrisinin altında kalan alan "area under the ROC curve" tanısı olan gruptan seçkisiz olarak alınacak bir katılımcının test sonucuna bağlı olarak hasta olarak sınıflandırılma olasılığını gösterir.

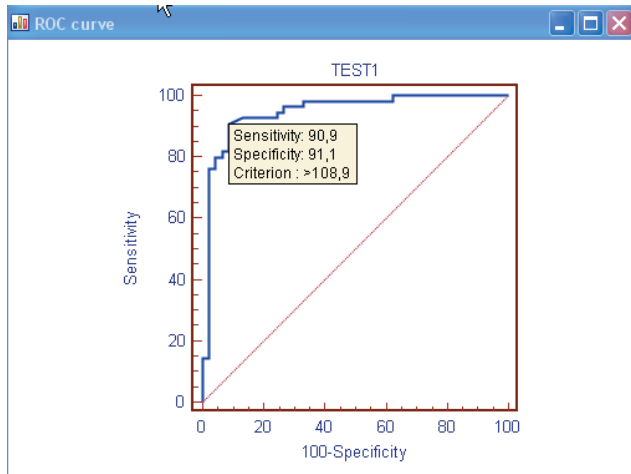
Aşağıdaki örnek analiz sonucuna göre, bu oran belirlenen kesim noktası için % 94.7'dir. ROC eğrisinin altında kalan alan için ulaşılan değer .5 ise bu teste dayanılarak yapılan sınıflamaların yazı tura atmaktan farklı olmadığı kabul edilir. Bu değer 1 olması, testin tanı alan ve almayan bireyleri mutlak bir doğrulukla ayırdığını gösterir.



Analiz sonucu elde ulaşılan kesim noktasının önünde “*” işareti bulunur. Bu örnek için 108.9 ve üzeri puanlar katılımcının tanı alması gerektiğine işaret etmektedir. Diyalog kutusunda “Full ROC

report” işaretlenmişse, Duyarlılık (Sens.), Özgüllük (Spect.), Pozitif Yordayıcı Değer (+LR) ve Negatif Yordayıcı Değer (-LR) her bir olası Eşik Değer için hesaplanmış olarak gelir.

>107,7	92,73	82,4 - 97,9	82,22	67,9 - 92,0	5,22	0,088
>107,9	92,73	82,4 - 97,9	84,44	70,5 - 93,5	5,96	0,086
>108,3	92,73	82,4 - 97,9	86,67	73,2 - 94,9	6,95	0,084
>108,9 *	90,91	80,0 - 96,9	91,11	78,8 - 97,5	10,23	0,10
>110	89,09	77,7 - 95,9	91,11	78,8 - 97,5	10,02	0,12
>110,4	87,27	75,5 - 94,7	91,11	78,8 - 97,5	9,82	0,14
>110,5	83,64	71,2 - 92,2	91,11	78,8 - 97,5	9,41	0,18



Grafik:

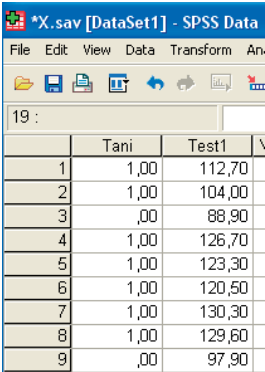
ROC eğrisinin ifadesinde dikey eksen duyarlık puanlarını 100 üzerinden, yatay eksen ise yanlış tanı oranını 100-Özgüllük üzerinden ifade eder. Eğri, diagonal eksenden uzaklaştıkça testin tanı açısından doğruluğunun arttığı gözlenir. Test için ulaşılan kesim noktası ve o noktadaki duyarlılık/özgüllük değerleri grafik üzerindeki kutucukta verilir.

SPSS 15.0

MedCalc ve SPSS programlarındaki uygulama ve yorumlamaların oldukça benzer olduğu gözönünde bulundurularak SPSS paket kullanımıyla ilgili açıklamalar sınırlı tutulmuştur.

Veri girişi:

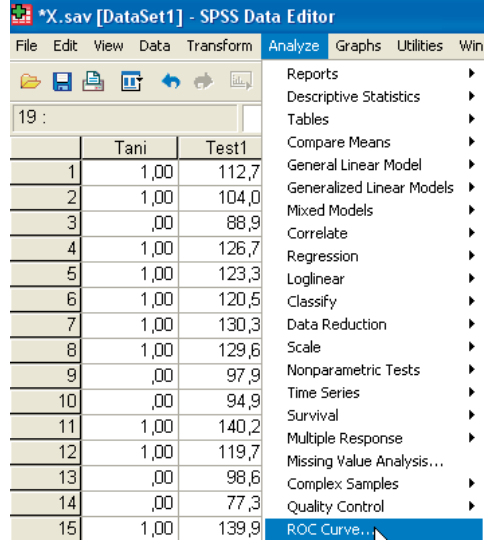
Veri girişi MedCalc'da olduğu gibi 2 sütun halinde yapılır.



	Tani	Test1
1	1,00	112,70
2	1,00	104,00
3	,00	88,90
4	1,00	126,70
5	1,00	123,30
6	1,00	120,50
7	1,00	130,30
8	1,00	129,60
9	,00	97,90

Analyze

..... ROC curves



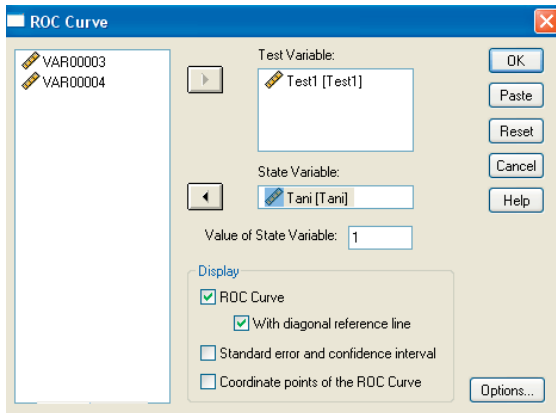
Komutlar:

Aşağıda gösterilen komut sistemi takip edilerek diyalog kutusu açılır.

Açılan diyalog kutusunda doldurulması gereken alanlar aşağıda kısaca tanımlanmıştır.

Test Variable: Üzerinde çalıştığımız değişkeni soldaki kutucuktan seçerek buraya girin.

State Variable: Kategorik değişkeninizi soldaki kutucuktan seçerek buraya girin.



Value of State Select: Katagorik değişken için tanıyı temsil eden sayısal değeri bu kutucuğa girin. Bu örnekte tanı alma durumu "1" ile tanımlanmıştır.

Display: Analiz sonuçları içinde görmek istediğiniz parametreleri "Display" bölümünün altında yer alan kutucuklarda işaretleyin. ROC eğrisinin diagonal eksene göre yerleşimini göre bilmek için "with diagonal reference line" seçeneğini işaretlemeniz gerekir.

Sonuçlar:

Sonuç sayfasındaki ilk bölüm tanı için örneklemdaki gözlem sayılarını verir.

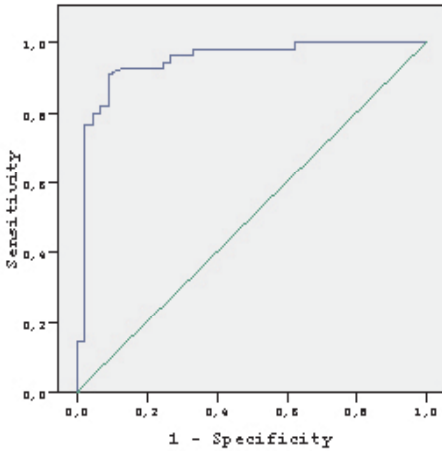
Case Processing Summary

Tanı	Valid N (listwise)
Positive ^a	55
Negative	45

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1,00.

ROC Curve



Area Under the Curve

Test Result Variable(s): Test1

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,947	,024	,000	,900	,994

The test result variable(s): Test1 has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

İkinci bölümde, ilgili ROC eğrisi verilmiştir. MedCalc'tan farklı olarak, duyarlık puanları dikey ekseninde 0 ile 1 aralığında, yanlış tanı oranı ise yatay ekseninde 1 - Özgünlük olarak ifade edilmektedir.

Üçüncü bölümde ROC eğrisinin altında kalan alana ilişkin bilgiler aktarılmıştır.

ROC eğrisinin altında kalan alana ilişkin detaylı tabloya ulaşabilmek için diyalog kutusundan "standart error and confidence interval" seçeneğinin işaretlenmesi gerekir. Bu seçenek kullanılmadığı durumda program sadece alan bilgisini hesaplar.

Area Under the Curve

Test Result Variable(s): Test1

Area
,947

The test result variable(s): Test1 has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Son bölümde, farklı eşik değerler için duyarlılık ve 1 - Özgünlük değerleri aktarılır. Bu bilgiye ulaşabilmesi için "Display" seçeneğinden "Coordinates of the Curve" kutucuğunun işaretlenmesi gerekir. MadCalc'tan farklı olarak, SPSS'te duyarlık/özgünlük değerleri ayrı ayrı verilmemekte ve 1 üzerinden rapor edilmektedir.

Coordinates of the Curve

Test Result Variable(s): Test1

Positive if Greater Than or Equal To ^a	Sensitivity	1 - Specificity
76,3000	1,000	1,000
77,3500	1,000	,978
80,2000	1,000	,956
83,5000	1,000	,933
84,1000	1,000	,911
84,9000	1,000	,889
85,8000	1,000	,867
86,1000	1,000	,844
86,3000	1,000	,822

SPSS çıktısında hesaplanan kesim noktası özel olarak işaretlenmemiştir. Araştırmacının en yüksek duyarlılık ve en düşük 1 - Özgüllük değerini temel alarak kesim noktasını kendi belirlemesi gerekmektedir. Örnek çalışma için kesim noktası 108.60 olarak hesaplanmıştır.

106,1500	,927	,222
107,3000	,927	,200
107,8000	,927	,178
108,1000	,927	,156
108,6000	,927	,133
109,4500	,909	,089
110,2000	,891	,089
110,4500	,873	,089
110,6000	,836	,089

SPSS paket programında Pozitif Yordayıcı Değer ve Negatif Yordayıcı Değerlerin elle hesaplanması gerekmektedir.

Genel izlenim olarak, MedCalc paket programının ROC analizi açısından daha fazla seçenek sunduğu, daha detaylı bir analiz dökümü verdiği ve eşik değer hesaplamalarında daha yakın değerler kullandığı görülmektedir.