

Kriket Maçları 1997-2005 Sezonları Analizi

Saida Meybaliyeva*

Özet

Kriket, on birer kişilik iki takım arasında, sopa ve top yardımıyla, merkezinde 20 metre (22 yard) uzunluğunda dikdörtgen bir saha bulunan ovalimsi bir alanda oynanan takım sporudur. Dikdörtgen sahanın bir ucunda yer alan atıcı durumundaki takımın oyuncusu topu, vurucu konumundaki takım oyuncusunun bulunduğu diğer uca doğru fırlatarak vurucunun arkasında bulunan kale benzeri düzeneğe (wicket) değdirmeye, vurucu ise gelen bu topun wicket'a değmesini engellemek için elindeki sopa yardımıyla topa vurmaya çalışır. Başarılı bir vuruşun ardından atıcı takım oyuncuları topu yakalamaya çalışırken, sahada yer alan vurucu da karşısındaki wicket'a doğru koşar. Yaptığımız analizde kazanılan skorlar arasında anlamlı pozitif korelasyon olduğunu gördük. Normal olarak olması gereken gibi kaçırılan topla da skor arasında negatif korelasyon olduğunu bulduk. Zaman açısından yıllara göre oyunlardaki toplam skorun medyan değerine baktığımızda da gözeçarpan bir değişiklik olduğu söylenemez.

1 Giriş

Kriket 16'cı yüzyılda İngilterede çıkan ve oradan dünyaya yayılan bir takım oyunudur. Şuanda Türkiyede pek oynanılmasa da eski İngiliz sömürgesi olan ülkelerde(Hindistan,Pakistan,Bangladeş) yaygın olarak oynanılmakta ve popülerliğini korumaktadır. Ofansif taraf fırlattığı topla savunma yapan oyuncunun arkasındaki kaleyi hedef alır(*Bowlers*) ve savunma yapan kişi topu arkasındaki kaleye değdirmeyecek şekilde elindeki sopayla savunmak zorundadır(*Batsmen*). Her takımda 11 kişi bulunur, top eğer kaleye değmediyse savunan taraf topa, topu savuran tarafın oyuncuları ise kaleye doğru koşarlar.

Analizde kullanılacak data **Github** üzerinden **Tidytuesday** reposundan alınmıştır. Data 16 sütun ve 1237 satırdan oluşmaktadır. **Takım1** ve **Takım2** oyuncu takımların ülkelerini aynı şekilde **skor1** ve **skor2** uygun gelen skorları, **zaman** maçın oynandığı zaman durumunu, **wicket_team1** ve **2** kaleden kenara giden top sayısını(10 olması tüm topların kenara gittiği anlamına geliyor),**takımsaha1** ve **takımsaha2** maçın hangi takım sahasında oynandığını,**kazanan** sütunu hangi takımın kazandığını,**series** maçın oynandığı turnuvanı, **player_of_match** ve **player_of_match_team** maçın oyuncusunu ve bu oyuncunun hangi takımda olduğunu, **toss** ve **toss_decision** sütunları yazı-turanı hangi takımın kazandığını ve neyi seçtiğini, **match_date** ise maçın oynanma zamanını gösteriyor.

*20080742, [Github Repo](#)

1.1 Çalışmanın Amacı

Çalışmada veride olan sezonlar için genel analiz yapmak amaçlanıyor. Genel olarak zaman içerisinde hangi takımların yükseldiğini, düştüğünü veya stabil oynadığını, sezonlar için en iyi oyuncu, ortalama skorlarda yıllara göre bir değişiklik olup olmadığı gözlemlenecektir. Maçı hangi takımın kazanacağına en çok etki eden etkenlerin hangi sütunlar olduğu da regresyon ve korelasyon analizi ile incelenecektir. Tabii ki, elimizde olan data kapsamlı bir regresyon analizi için yetersizdir, çünkü oyuncuların genel bilgileri olmadan bir tahmin yapmak pek olası değildir, ancak yine de tahminler oyuncuların durumlarının değişmediğine göre yapılacaktır.

1.2 Literatür

Genel olarak ülkemizde ve yakın ülkelerde pek çok bilinmese de, kriket oyunu 18'ci yy'dan beri daha da yaygınlaşmıştır ve popülerliğini korumaktadır. İngilterede 2'ci meşhur spor olsa da Amerikan futbolu sayılan rugby'den 2 kat ve baseball'dan 1.3 kat daha fazla popülerdir, 5 yetiştikenden biri düzenli olarak kriket seyretmekte ve 10 yetiştikenden biri oyunlarla alakalı yazılar okumaktadır(Lemmer, 2008). Bu oyunun bu kadar çok popüler olmasına rağmen son 10 yıldır diğer sporlara nazaran daha çok araştırmalara konu olmaktadır. Amin ve Sharma (2014) araştırmasında oyunlar incelendiğinde maç yerinin ve oynandığı turnuvanın kazanma oranına göre takımlar için farklılığa yol açtığı bulunmuştur. Araştırmamızda bu etkenlere diğer etkenler de katılacak şekilde analiz yapılacaktır. Bunlar önemli etken olsalar da takım oyuncularının skora katkısı tartışılmaz şekilde en yüksek olandır. Ancak yine oyuncular da bir çok özelliğe göre farklılık göstermektedir. Genel olarak oyuncuların eski performanslarına göre onların diğer oyunda hangi düzeyde iyi veya kötü oynayacağını tahmin edilirken öznel tam olarak anlamlı olmayan sonuçlarla karşılaşılıyor(Iyer ve Sharda, 2009). Top savurma konusunda bile bir oyuncu hedef konusunda iyi olurken, bir diğeri hıza göre en iyi olabiliyor, yani takım isteklerine ve pozisyonlara göre iyi oyuncular vardır tam olarak en iyi oyuncu diye bir kavram yoktur ki, bu da kriket sporunun heyecanlı kılmaktadır(Mukherjee, 2014). Singh vd. (2015) araştırmasında kazanan takımı tahmin etmek için *lineer regresyon* kullanılmıştır. %91 güven düzeyinde kazanan takımın mevcut skor ve verimizde bulunan **wicket team** sütunlarıyla orantılı olduğu bulunmuştur. Genel olarak bakıldığında bu özelliğin diğer popüler olan sporlarda da bulunuyor. Hatta bir çok spor bunu korumak için gerekli önlemler alıyor. Formula 1'i son yıllarda Mercedes takımının daha çok para harcayarak domine etmesinden dolayı federasyonun dominasyonu kırmak için harcanan paraya bir kota konulmasını bir örnek olarak vere biliriz.

2 Veri

Tablo 1'da görüldüğü gibi maksimum skor 398 puan ve minimum ise 8 puandır. Skor 1 ortalaması 233 puan standart sapması ise 56 puan olarak gözüküyor. Medyan değeri ise 237 ile ortalamaya oldukça yakın. Aynı şekilde skor 2 değişkeninde ortalama 200 ve yine medyan değeri 203 ile ortalama ve medyan değerleri bir-birine çok yakındır. Kenara giden toplara bakarsak, tabii ki, maksimum olabilecek maksimum sayı 10 ve minimumsa 0'dır. Aynı şekilde

skorda olduđu gibi burda da ortalamalarla medyan deęerleri yakındır. Ortalamada çok puan olmasına rağmen 1-ci takımlarda kenara giden top ortalaması ve medyanı fazladır. Yani savunucular görevlerini yerine getirmektedirler. Kaleye giden topun da tahtaya isabet etmesi çok düşük ihtimaldir. Total skora bakacak olursak da yine en düşük skorlu maçta toplam 73, en yüksek skorlu maçta ise skorun 693 olduğunu görürüz. Total skor açısından da medyan değeri 441 puandır.

Tablo 1: Özet İstatistikler

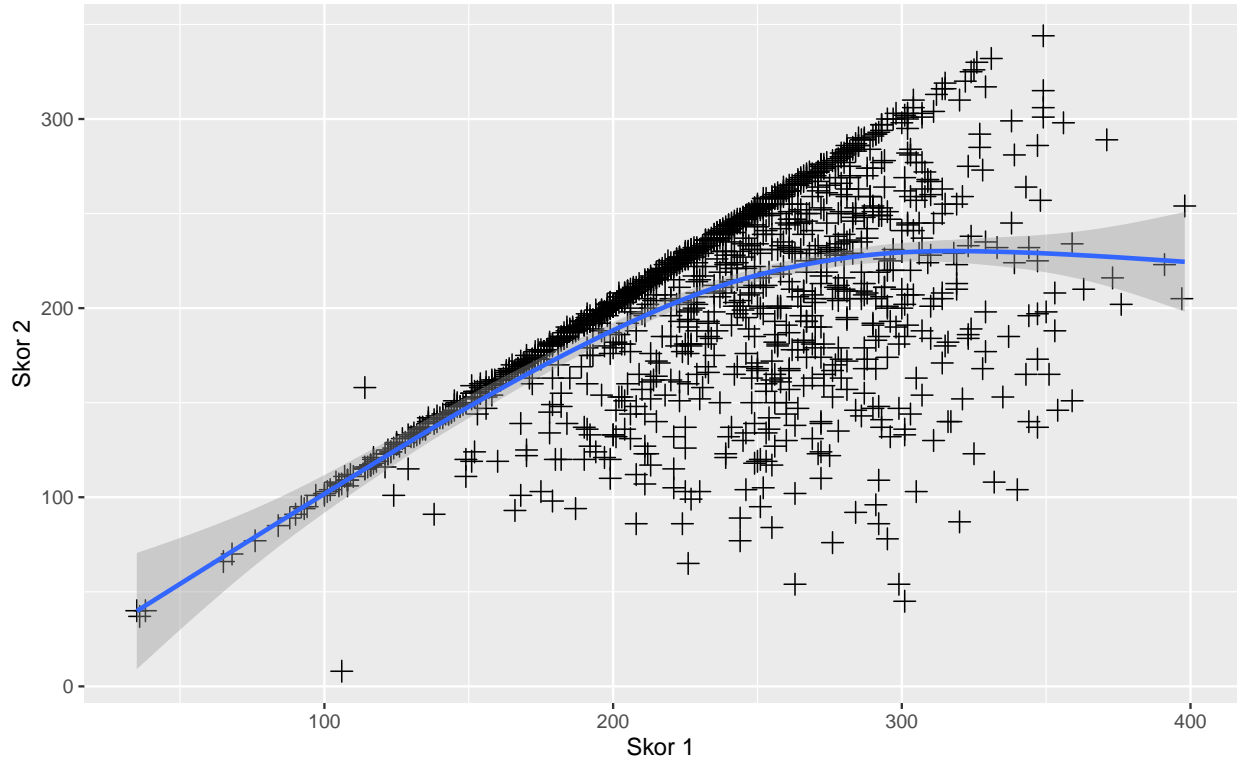
	Ortalama	Std.Sap	Min	Medyan	Mak
skor1	233.17	56.83	35.00	237.00	398.00
skor2	200.29	52.21	8.00	203.00	344.00
total_skor	433.46	96.90	73.00	441.00	693.00
wickets_team1	7.74	2.03	2.00	8.00	10.00
wickets_team2	6.76	3.14	0.00	7.00	10.00

Tablo 2: Shapiro-Wilk Normallik Testi

	Test İstatistikleri	P Değeri
W	0.9906976	5e-07

3 Yöntem ve Veri Analizi

Şekil 1’de skor 1 değişkeninin neden daha fazla ortalamaya sahip olduğunu görüyoruz. Ekseriyette fark açılan oyunlarda kazanan takımın skoru skor 1 sütununda yer almaktadır. Sol üst köşenin boş olmasından ve sağ alt köşede daha çok nokta olmasından bunu anlayabiliriz. Skoru 400’e yakın olan maçlar da genellikle skor 1dedir. Regresyon doğrusunun aynı düzeyde devam etmemesi de yüksek olan skorların skor1’de yazılmasından kaynaklanmaktadır.

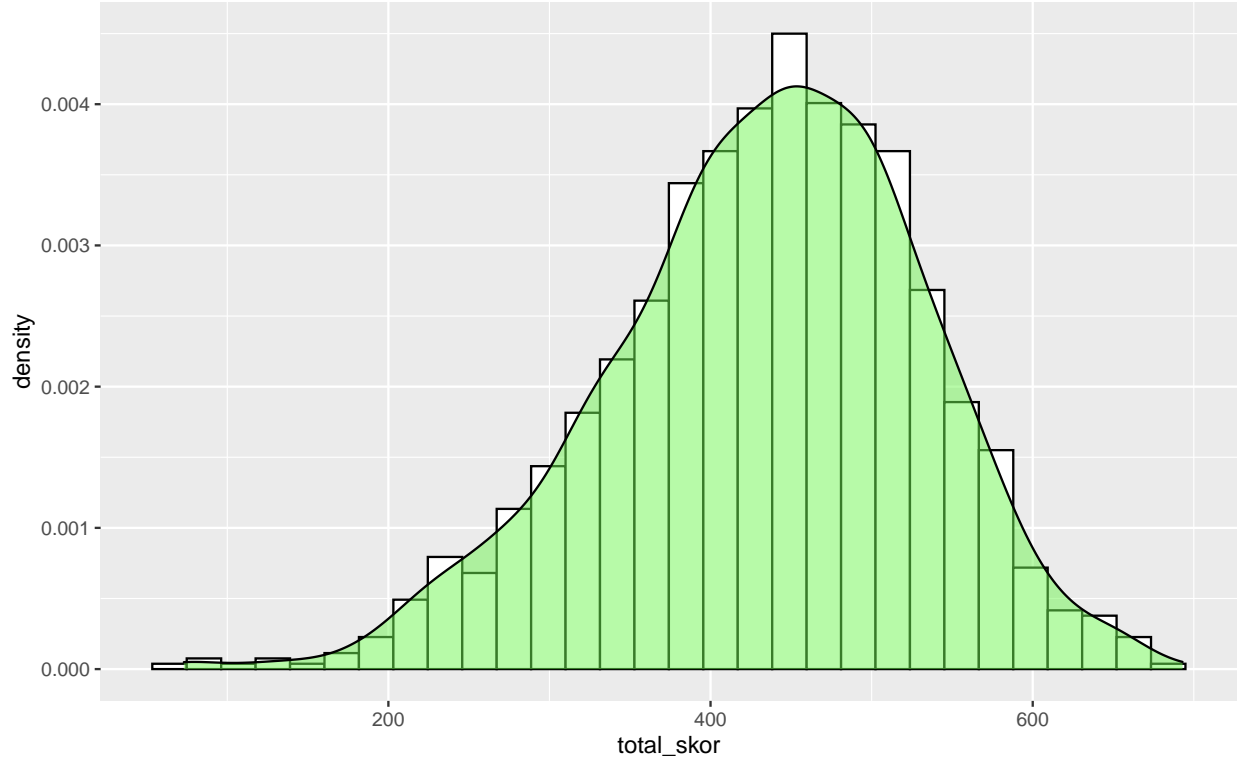


Şekil 1: Skor Saçılım

Şekil 2’deki yoğunluk grafiğine bakarsak en fazla oranla maçların toplamda 380-520 skorla bittiğini göre biliriz. Dağılım sola doğru çarpık olarak gözüküyor. Oyunda sayı 200’den az ve 640’dan fazla skor olanlar genelin azını teşkil etmektedir. Dağılım normal dağılıma benziyor. Tablo 3’deki Shapiro testinden de toplam skorun normal olarak dağıldığını görüyoruz.

Şekil 3’da yıllar bazında en yüksek ortalama skorun verinin toplandığı en son yıla denk geldiğini görüyoruz. En düşük skorlu yıl 2003 yılına aittir. Genel olarak ortalama skorlar pek farklılaşmamakla birlikte yıllar bazında da önemli bir çizgi göremiyoruz.

Tabii ki, kenara giden top sayısı arttıkça skor kazanmak da takım açısından zorlaşıyor. Skorla

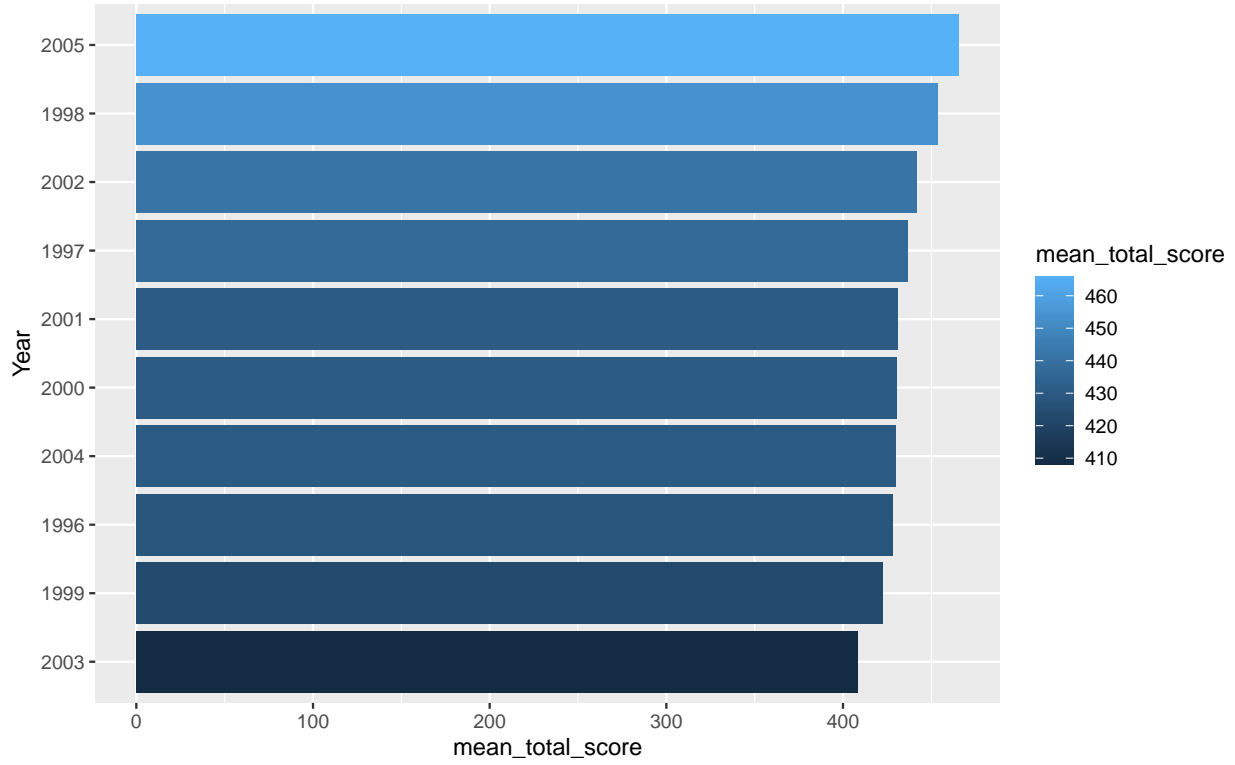


Şekil 2: Skor Histogramı

Tablo 3: Korelasyon Test Sonuçları

Değişken	Test_İstatistikleri	P_Değeri
skor1 vs. wickets_team1	-27.24329	0
skor1 vs. skor2	24.92879	0

kaçırılan top sayı arasında korelasyona baktığımızda istatistiksel açıdan anlamlı korelasyon katsayısını -0.61 olarak buluyoruz. Yani bu iki değişken arasında ters yönlü bir ilişki vardır.



Şekil 3: Yıllara göre Ortalama Skorlar

4 Sonuç

Sonuç olarak maçlardaki skorların eşit ağırlıklı yani normal olarak dağıldığını gördük. Shapiro normallik testiyle bunu bir daha doğruladık. Yıllar bazında ortalama skorlar açısından belirgin farklılık bulunmamaktadır.

Şekil 2’de de skorlar arasında bir aynı yönlü hareket olduğu görünüyor. Korelasyon testiyle de bu iki değişken arasında aynı anlamlı 0.58 oranlı korelasyon katsayısının olduğunu görüyoruz. Yani bir takımın yüksek(düşük) skor alması diğer takımın da aynı şekilde skorlar üretmesine tahrik ediyor. Ayrıca kaçırılan toplarla skor arasında negatif korelasyon da var. Bu iki korelasyonun P değerlerine bakılırsa istatistiksel olarak da anlamlı olduğunu görüyoruz.

Oyunların total skor açısından dağılımına baktığımızda genel olarak en çok yüzdeliği alan intervalin 400-500 sayı arasında ve median, ortalama değerler de bu intervalle denk geliyor. Maksimum değere nazaran minimum değer olan 73 ve üstünün çok az olasılıkla gerçekleştiğini gördük.

Elimizdeki sayısal değerler az olduğundan kapsamlı istatistiki analiz gerçekleştirmek zor, detaylı süreler oyuncular hakkında özel bilgiler veriye ilave olunursa daha iyi sonuçlar alınabilir. Maç içi zamana göre veriler olursa, kaçırılan toplardan ve oyunculara göre oyunun skoru lineer regresyonla tahmin edilebilir.

5 Kaynakça

- Amin, G. R. ve Sharma, S. K. (2014). Cricket team selection using data envelopment analysis. *European journal of sport science*, 14(sup1), S369-S376.
- Iyer, S. R. ve Sharda, R. (2009). Prediction of athletes performance using neural networks: An application in cricket team selection. *Expert Systems with Applications*, 36(3), 5510-5522.
- Lemmer, H. H. (2008). An analysis of players' performances in the first cricket Twenty20 world cup series. *South African Journal for Research in Sport, Physical Education and Recreation*, 30(2), 71-77.
- Mukherjee, S. (2014). Quantifying individual performance in Cricket—A network analysis of Batsmen and Bowlers. *Physica A: Statistical Mechanics and its Applications*, 393, 624-637.
- Singh, T., Singla, V. ve Bhatia, P. (2015). Score and winning prediction in cricket through data mining. *2015 international conference on soft computing techniques and implementations (ICSCTI)* içinde (ss. 60-66). IEEE.