MODEL SEÇİMİNDE BAYESİAN YAKLAŞIM

*Dr. İbrahim DUYAR**

Abstrak

Bu çalışmanın temel amacı Bayesian yaklaşımını sosyal bilim araştırmaları model seçimi içeriğinde tanıtmak ve tartışmaktadır. Bayesian yaklaşım, model seçiminde kullanılan klasik P-değerlerine dayalı likelihood-ratio testleri (LRT) ile karşılaştırılmaktadır. Klasik LRT testlerinin model seçmedeki güçlükleri incelenerek Bayesian yaklaşımın bu güçlükleri nasıl giderdiği vurgulanmaktadır.

GİRİŞ

P-değerleri ve bu değerlere bağlı manidarlık testleri sosyal bilimlerde istatiksel çıkarsamada yaygın olarak kullanılagelmektedir. Bununla birlikte nicel tekniklerle çalışan sosyal bilim araştırmacılarının son 15 yıldır bazı pratik zorluklar nedaniyle analizlerinde P-değerlerine daha az yer verir olmuşlardır.

P-değerleri ile ilgili güçlüklerden en belirgin olanı, P-değerlerini büyük örneklemlerde, teorik temelleri kabul görmüş null hipotezini red etme eğiliminde olmasıdır. Bu problem, özellikle büyük örneklemlerle çalışan sosyal bilim araştırmalarında "ikinci tür" istatiksel hataları yaygınlaştırmıştır.

P-değerlerinin model seçimiyle ilgili diğer iki güçlüğü daha az belirgin olmakla birlikte, uygulamada daha yaygın karşılananıdır. Bu güçlükler özellikle veri analizinin ilk aşamalarında ve birden fazla model göz önüne alındığında meydana gelmektedir. Bu durum birçok kontrol değişkeni arasından hangilerinin modele dahil edilmesi gerektiği konusunda karar verirken gözlenir. P değerleri dikkate alınarak yapılan bu uygulamalar ya (a) tüm değişkenlerin t- değerlerinin gözden geçirilmesi ve en az manidar olanlarının ayıklanması ile ya da (b) daha formal bir yol olan aşamalı (stepwise) regrasyon şeklindedir.

---

* Eğitim Bilimleri Fakültesi Araştırma Görevlisi

Bu zorluklardan birincisi Miller'in (1984;1990) da belirttiği gibi p-değerine dayalı olarak geniş bir değişkenler setinden seçilen modelin yalnızca iki model gözönünde alındığında da aynı yorumu veriyor olmasıdır. Fredman (1983), Freedman, Navidi ve Peters (1988) P-değerlerinin model seçimini izleyen kullanımının ileri derecede yanıltıcı sonuçlar doğurabileceği konusunda benzer endişeleri dile getirmektedir.

İkinci zorluk, aynı veri ile ilgili değişik modellerin tümünün kabul edilebilir olmasına karşılık, bu modellerin herbirinin, üzerinde durulan konu ile ilgili farklı sonuçlara götürüyor olmasıdır. Bunu en çarpıcı örneği Kass ve Raftery'nin (1995) sosyal tabakalaşma çalışmasında gözlenmiştir. Bu çalışmada, yalnızca bir modelin dikkate alındığı ve yorumun buna bağlı olarak yapıldığı standard model seçme yönteminin, gözönüne alınan çocuklarla ilgili belirsizlikleri ve buna bağlı olarak da model formu ile ilgili belirsizlikleri gözönüne almadığı sonucuna gidilmiştir.

1980'lerin ilk yıllarında bazı sosyal bilimciler (Fienberg and Maso, 1979; Hout, 1983, 1984; Grusky and Hauser, 1984) teorik temellerle uyuşmayan sonuçların elde edildiği durumlarda, model seçimini P-değerleri yerine teorik temellere ve veri ile model arasındaki uyumsuzlukların informal değerlendirilmesine dayandırmaya başlamıştır. İlk kez Schwartz (1978) ve daha sonra Rafterry (1986) Bayesian yaklaşımı ve özellikle Bayesian Information Criterian'i (BIC) hipotez testi ve model seçiminde P değerine dayanan testlere alternatif olarak sosyal bilimcilerin dikkatine sunmuşlardır.

Bayesian yaklaşım, P-değerine dayanan testlerin yukarıda açıklanan güçlüklerini ortadan kaldıran özelliği nedeniyle kısa zamanda teori geliştirme ve test etme amaçlı sosyal bilim araştırmalarında geniş kullanım görmeye başlamıştır. Öncelikle Bayesian yaklaşım, P-değerlerinin desteklenmesine karşılık teorik temellere bağlı "sezgisel genel kabulun" geçerliliğini sağlamaktadır. (Raftery, 1994). Ayrıca Bayesian yaklaşım model seçimi ve model belirsizliğinin 1986 yılında Raftery (1986) tarafından pür bir model seçme ölçütü olarak kullanılmış ve bu tarihten sonra yaygın bir şekilde kullanılır olmuştur.

İleriki sayfalarda P-değerli likelihood-ratio testlerinde karşılaşılan bazı zorluklar ayrıntıları ile incelenmiştir.

<u>P-Değerleri İle İlgili Güçlükler</u>

Hipotez test etme ile ilgili standard yaklaşım yalnızca alternatif, H_1, ve bu alternatif hipotez içinde yuvalanmış (nested) bir nul, H_0, hipotezi olduğunu varsayar. Alternatif hipotez bilinmeyen sayıdaki parametrelerden d_i oluşan modeli $Ø=(Ø_1,...,Ød_1)$ ile gösterilir. Null hipotezi de aynı olasılık modeline $Ø_{,gi}(Ø)=0$ $(i=1,...\surd)$ $\surd$ sınırlaması yapılarak gösterilebilir.

Bir t-test istatiği eldeki verilerden, D, hesaplanır ve gözlem değeri t (D) ile gösterilir. Eğer t(D) beklenenden daha yüksek ise H_0 hipotezi H1 lehine red edilir. Bu prosüdür bir manidarlık seçilerek başlar ve H_0'nin T-tablo değer olasılığının t(D)'den (α değerinden) küçük ya da eşit olması ile red edilmesi ile uygulamaya konur. Daha formal bir ifade ile H_0 aşağıdaki durum sözkonusu ise red edilir.

$$P = Pr\,[T \geq t\,(D) \mid H0] < \alpha \qquad (1)$$

Araştırmalarda P-değerinin büyüklüğü H_0'ne karşı katının gücü olarak rapor edilir.

Bu yaklaşımın temelleri çoğu zaman eleştirimeden kullanılır. Oysa dikkat edilmesi gerekli bazı önemli noktalar bulunmaktadır. Bunlardan biri manidarlık düzeyinin belirlenmesi zorunluluğudur. Uygulamada Fisher'in çalışmalarında önerilen .05 ya da .01 düzeyleri yaygın olarak kullanılmaktadır. Manidarlık düzeyi belirlenirken alternatif hipoteze karşı olan testin gücünün belirlenmesi ve güç ile manidarlık düzeyi arasında uygun bir denge kurulması gereklidir. Bu dengeyi yakalamanın en kesin yolu büyük örneklemeler için α'i olabildiğince küçük almaktır. Uygulamada bu yol genellikle tercih edilmemektedir.

İkinci nokta, tüm standart hipotez testinin yalnızca iki modele dayandırılacağı varsayımdır. Bu varsayım, birçok olası kontrol değişkenlerinin gözönüne alınmasını gerektiren eğitim ve diğer sosyal bilimlerinin gözönüne alınmasını gerektirmekte ve eğitim ve diğer sosyal bilim araştırmalarındaki durumla uyuşmamaktadır. Uygulamada seçim değişkenler arasında yapılmaktadır ve her seçim ayrı bir modele karşılık gelmektedir. Her olası p değişkenin eklenmesi durumunda 2^p sayısında aday model bulunacaktır. Örneğin, 15 değişkenin içerildiği bir durumda 2^{15} = 32.768 aday modelden söz edilir olmaktadır. Bu daha fazla değişken olması durumunda pratik olmayacak büyüklüklere ulaşmaktadır.

Üçüncü nokta, (1) numaralı formül tanımı Ho'ne dayanmaktadır. H_0'nin red edilmesi durumunda ancak H_1 kabul edilmektedir. Oysa, gerçekte birden çok alternatif hipotez bulunmaktadır. Örneğin, regrasyonda alternatif hipotezler hataların normal dağılımı ya da geniş bir bağımsız değişkenler seti olabilir. Logaritmik doğrusal (log-linear) modellemede ya aşırı dağılma ya da ek marjinlerin uyması (fit) söz konusu olabilir. Uygulamada bunların gerçek bir güçlük oluşturmadığı ve araştırmacıların birkaç alternatif hipotezden verilerine en iyi uyanı (best fit) seçtiği ileri sürülmektedir. Fakat bu durumda da H1'in sabit(fix) olma temel varsayımı ihlal edilmiş olur.

Berger ve Sellke'ye göre (1987) göre buna en iyi çözüm belli bir alternatif hipoteze karşılık güçlü bir T-test değeri kullanmaktadır. Burunla birlikte, bu çözümde tümüyle yeterli olmayabilir. Örneğin, H1'den başka diğer alternatif hipotezler T-değerlerinin büyümesine ve H0'nin red edilmesine yol açar. Regrasyonda belli bir bağımsız değişkenin etkisinin olup olmadığını test etmek istediğimizi farzedelim. T-değerlerinin kullanıldığı bu durumda uzun kuyruklu normal olmayan bir hata dağılımı, etkisi büyük olan normal dışı (outlier) gözlemleri sıkça ortaya çıkabilir ve böylece T-değerlerinin büyük olmasına ve H0'nin red edilmesine yol açar. H0'nin red edildiği bu durumlarda normal olmayan hataların mı yoksa ek bağımsız değişkenlerin mi kabul edileceği araştırıcının önceden vermiş olduğu karara göre belirlenir.

<u>Büyük Örnekler</u>

Sosyal araştırmaların birçoğu bir sosyal olgunun belirli yönleri ile ilgili özelliklerini betimlemektedir. Uygulamada betimleme, genellikle üzerinde çalışılan olguyu temsil eden modeller ile eşanlamlı kullanmaktadır. Bununla birlikte likelihood-ratio testleri ve diğer manidarlık testleri model ile sosyal olgu arasındaki uyumsuzlukları (discrepancies) ölçmek üzere geliştirilmişlerdir. Diğer bir deyişle bu testler eldeki mevcut verinin modele ne ölçüde uyduğunu ya da uymadığını test eder. Model kabul edilebilir olsa bile bu uyumsuzluklar tanım gereği az da olsa mutlaka bulunacaktır. Yeterince büyük örneklemlerle LRT bu uyumsuzlukları bulacak ve iyi olmasına rağmen modeli red edecektir.

Örneğin, M_0'nun M_1 lehine red edildiğini farz edelim. Mo'nun reddinin M1'in daha iyi betimleme sağladığı şeklinde yorumlamak doğru değildir. Çünkü temel olarak yapılan şey, modellerin karşılaştırılması olmayıp, bu modellerden biri ile mevcut veri arasındaki marjinal uyumsuzlukların belirlenmesidir. Raftery (1994) araştırıcıların bu konuda yanılgıya düşüp, likelihood-ratio testleri ile modellerin karşılaştırılması yapıyorlarmış şeklinde yanlış kullanımlara düştüklerini ileri sürmektedir. Grusky ve Hauser (1984) de özellikle büyük örneklemli çalışmalarda bu yanlış kullanımların yanlış sonuçlara ulaştıracağını kendi çalışmalarında kullandıkları büyük örneklem ile (n=113.556) uygulamalı olarak göstermektedirler.

Birçok Aday Bağımsız Değişken Durumu

Birçok sosyal bilim araştırması bağımlı ve bağımsız değişken arasındaki ilişkilerin belirlenmesi amacıyla yapılır. Bu araştırmalar hangi bağımsız değişkenlerin dikkate alınacağını kararlaştırmak durumundadırlar. Kuşkusuz hangi değişkenlerin seçileceğine teori rehberlik eder. Bununla birlikte teoriden çıkarılan değişkenler çoğu zaman uzun listeler oluşturur, bunlardan bir kısmı da oldukça yüzeysel teorik temellere sahiptir. Diğer bir deyişle, çoğu zaman teori ne yazık ki, üzerinde çalışma yapılacak sosyal olgu ile ilgilil kesin ve sınırlı (parsimonious) bir değişkenler listesi sunmaz. Biship, Fienberg ve Holland (1975) kesin olmayan uzun değişkenler listesinin, hiç etkisi olamayan ya da ihmal edilebilir etkileri olan değişkenleri içermesi nedeniyle üzerinde çalışılan parametere tahminlerinin doğruluğunu azalttığı ileri sürmektedir.

Hangi değişkenlerin secileceği konusunda genel kabul gören ve uygulama bulan bazı teknikler vardır. Bunlardan biri T-istatiklerin gözden geçirilmesidir. T-değerleri küçük olan değişkenler ayıklanır. Diğer bir yol aşamalı regrasyondur. Bu teknikte: (a) null modelinden başlayarak her defasında bir değişken ayıklayarak (backwards selection), ya da (c) her ikisinin karışımı uygulanabilir. Mallows'un C_p ve R^2'in maksimize edilmesini diğer değişken seçim teknikleridir. (Miller, 1990).

Bu tekniklerin ortak yanı, onların birçok olasılıktan yalnızca bir model seçmeleri ve daha sonra başından beri yalnızca bir tek modeli inceliyormuş gibi davranmalarıdır. Bu tür bir uygulamanın yanıltıcı olduğu birçok yazar (Freedman, 1983, Freedman, Navidi ve Peters, 1988, Fenech, ve Westfall 1988, Miller 1984;1990) tarafından dile getirilmiştir. Bu yazarlar birçok sayıdaki modelin seçim yapmanın, diğer şartlar bir kenara bırakılsa bile tek başına manidar değişkenleri elde etme olasılığı yükselttiği görüşünde birleşmektedirler.

Raftery (1994) ve Freedman (1983) çalışmalarında birer örnek ile model seçiminde aşamalı regrasyonun özellikle çok değişken gözden geçirilmesi durumunda nasıl yanıltıcı olduğunu göstermişlerdir. Freedman aşamalı regrasyonun ikinci elemesinde bağımsız ve bağımlı değişkenler arasındaki ilişkilerin "duman ile duman" arasındaki ilişki kadar kesin olduğunu ileri sürmektedir.

Model Belirsizliği

Birden çok model dikkate alındığında bir kaçının birden etkili verilere eşit derecede uyduğu görülebilir. Aynı şekilde değişik model seçme yöntemleri kullanarak, aynı konu ile ilgili değişik modellere ulaşmak mümkündür. Bu herbiri kendi bazında savunabilecek modellerin aynı konu ile ilgili olarak değişik sonuçlara ulaştırması demektir.

Bu konu ile ilgili olarak alanyazında üç değişik uygulamaya gidilmiştir. Birincisi alanyazından belli bir teori seçip sonuçları yalnızca bu model çercevesinde yorumlamaktır. İkinci yol, analiz sonuçlarını uyan her modele ilişkilendirerek sunmaktır. Bu yol tümüyle kabul edilebilir olmasa da birinciye nazaran daha ileri bir düzeyde seçimdir. Üçüncü bir yol ise, sonuçlara giderken modelin belirsizliğini açık bir şekilde kabul etmektedir. Bu yol daha sonra geniş olarak tartışılacaktır.

Null Hipotezi İçin Kanıt

Standard manidarlık testleri ya null hipotezinin red edilmesi ya da red edilmesinde başarısız olmaya izin verir. Bununla birlikte standard manidarlık testleri null hipotezi için herhangi bir kanıt ölçümü sağlamazlar. Bazen sosyal teoriler inceleme konusu yapılan bir konunun değişik gruplarda aynı olduğunu gösterebilir. Bu durumlarda inceleme konusu null hipotezidir. Fakat stardard manidarlık testleri bu tür hipotezlerin doğrudan test edilmesine imkan tanımazlar.

Özetlemek gerekirse, standard bir test verinin destekleyip null hipotezini desteklemediği konusunda hiç bir bildirimde bulunmadan yalnızca red ettiği ya da etmede başarısız olduğunu gösterirler. Oysa Raftery'nin (1994) de belirttiği gibi bir test null hipotezini ya yeterli veri bulunmadığından ya da verinin onu desteklediğinden red etmekte başarısız olur. İşte standard testler null hipotezini red etmede başarısızlığın hangi nedenden kaynaklandığını göstermezler.

Hipotez Test Etmede Bayesian Yaklaşım

Bayesian yordama bir modemin olasılık açısından bilinmeyen bütün parametrelerini içerir. Bilinmeyen bütün parametreler random değişkenler olarak değerlendirilir. Bu nedenle Bayesian istatistik doğrudan temel olasılık teorisine dayanmaktadır.

Bilinmeyen parametrelerin vektörü $Ø=(Ø_1,...Ø_d)$ ile gösterilen D verisi için bir olasılık modeli düşünelim. Veriler gözlenmeden önce Ø'nin belirsizliği hakkındaki düşüncemiz önceki (prior) dağılım p(Ø) ile gösterilir. Olasılık modeli, Ø gerçek parametre veri (given) kabul edilerek ve D verisinin gözlenebilme olasılığı p(D|Ø) ile belirlenir.

Veri D gözlendikten sonra, yine D veri kabul edilerek, Bayes teoremini kullanark Ø'nin sonraki (posterior) dağılımı elde etmek için Ø hakkındaki düşüncelerimizi yeniden gözden geçiririz.

$$p\,(Ø|D)\ \alpha p\,(D|Ø)\ p\,(Ø) \qquad (2)$$

Sonuçta, sonraki dağılım, olasılık ve öncesi dağılımın çarpımları ile orantılıdır.

Sonraki dağılım p(Ø|D)Ø hakkında yordama yapmak için ihtiyaç duyulabilecek her türlü infarmosyonu içermektedir. Bu nedenle sorulabilecek tek soru, bu infarmosyonun nasıl en iyi şekilde özetlenmesi gerektiğidir. Uygulamada en çok başvurulan yol Ø'nin elemanları hakkında derinlemesine çalışılmalıdır. Örneğin, önceki dağılımın elemanlarından olan Ø'nın, diyelimki Ø1 hakkında yordamı şu şekilde yapılabilir:

$$p(\varnothing 1|D) = \partial p(\varnothing|D)\, d\varnothing 2 d\varnothing 3 \ldots d\varnothing_d \quad (3)$$

Bu univarite dağılımında $\varnothing_1$ hakkında yordama yapmak için her türlü informasyon bulunmaktadır. Örneğin, bu informasyonu özetlemek için %95 Bayesian güven aralığını tanımayan .025 ve .975 yüzdelik çeyrekler, sonraki standard dağılım, sonraki (aritmetik) ortalama, sonraki mod gibi göstergeler sıkça kullanılmaktadır.

Bayesian yordama araştırmacı tarafından subjektif olarak belirlenen sonraki sapma p(Ø)'ı temel alması nedeniyle eleştirilmektedir. Bununla birlikte subjektif belirleme büyük örneklemlerle hiç bir etkide bulunamamaktadır. Örneklem büyüklüğünün *n* olduğu bir durumnda sonraki ortalama ve varyansa etkisi, toplamın (1/*n*)'i kadar olmaktadır. (Raftery, 1994). Bu nedenle büyük örneklem durumlarında Bayesian ve maksimum olasılık (likelihood) yöntemlerinde aynı sonuçlar elde edilir. Fakat her iki yöntem küçük örneklemlerde farklı sonuçlar ortaya çıkarmaktadır. Bayesian yordama ile ilgili teknik ve uygulamalı temeller Press (1989), Lee (1989) Bernardo ve Smith (1994) gibi yazarlarca ayrıntılı olarak dile getirilmiştir.

Bayes Faktör

P değerine dayanan olasılık testlerinin eleştirildiği kısımda, model seçiminde amacın veri ile model arasındaki uyuşmazlığın tesbit edilmesi değil, modellerin karşılaştırılması olduğu vurgulanmıştı. Model seçimiyle ilgili kısmı daha net bir şekilde soru şeklinde belirtecek olursak veride yansıyan sosyal olgunun temel özelliklerini hangi model daha iyi betimler? Daha da öz bir biçimde ifade edilecek olursa: Mevcut veriler gözönüne alındığında farklı modellerde, örneğin M_0 ve M_1, hangisi gerçek modeldir?

Bu soru Mo modelinin M1 modeline karşı olan olan sonraki uyumsuzluklarını (odds) hesaplamakla bulunabilir.

$$B = \frac{\text{Olasılık } [M_0 \text{ gerçek modeldir}]}{\text{Olasılık } [M_1 \text{ gerçek modeldir}]} \quad (4)$$

Bayes faktor B, olasılık temeline dayanan, Bayesian yaklaşımın istatiksel yordama göstergesidir. Bayes faktörü hesaplamak için model ve parametreleri ile ilgili olarak mevcut verileri belirlemesi gerekir. Daha önce belirtildiği üzere büyük örneklemlerde mevcut verilerin sonuçlar üzerindeki etkisi ihmal edilecek büyüklüktedir.

Formül 4'te verilen konuyu daha ayrıntılı incelemek için Ø1 ve Ø2 parametreleri olan M_1 veM_2 gibi iki rakip modelden hangisinin mevcut veriye daha iyi uyduğunun test etmek istediğimizi düşünelim. Bu modeller yuvalanmış (nested) ya da yuvalanmamış olabilir. Bayes teoremi yardımıyla m_1 modelinin doğru model olduğunu aşağıdaki formül ile sınamak mümkündür.

$$p(D|M1) = \frac{P(D|M1)p(M1)}{p(D|M1)p(M1)+p(D|M2)p(M2)} \qquad (5)$$

Formülde $p(D|M_k)$ M_k verisine bağlı olarak marjinal olasılık, ve p (Mk) $M_k(k=1,2)$ modelinin önçeki olasılığıdır. Aynı açıklama p(M2/D) için de geçerlidir ve p(M1/D)+p(M2|D)=1 dir.

Mevcut verinin M_1 modeline karşı M_2 modelini desteklemesi M2'nin M'1'e karşı sonraki uyuşmazlıkların hesaplanması ie ölçülebilir. Bu amaçla formül 5 aşağıdaki gibi yeniden yazılabilir.

$$\frac{p(M2|D)}{p(M1|D)} = \frac{[p(D)|M2)]}{[p(D|M1)]} \times \frac{[p(M2)]}{[p(M1)]} \qquad (6)$$

Eşitliğin sağ tarafındaki birinci işlem, M_2 modelinin M_1 modeline karşı *Bayes faktörüdür* ve iki modelin birleştirilmiş olasılık oranını verir. Yine eşitliğin sağ tarafındaki ikinci işlem ise önceki uyumsuzlukları (prior odds) gösterir. Bu olumsuzluklar modellerden herhangi birine kredi verilmediğini gösterecek şekilde 1'e eşittir. Diğer bir deyişle, p (M_1)=p (M_2)=1/2.

Formül 6 kelimelerle şu şekilde ifade edilebilir:

(Önceki Olumsuzluklar)=(Bayes faktor)x(Önceki Olumsuzluklar) (7)

Yukarıdaki açıklama hatırlanacak olursa, Bayes faktor, ancak önceki olumsuzlukların 1'e eşit olması halinde sonraki olumsuzluklara eşit olacaktır.

M_1 ve M_2 modellerinin seçiminde Bayes faktor, $B_{21}>1$ ise M_2 modelinin M_1 modeline tercihi; $B_{21}<1$ durumunda ise M1 modelinin M_2 modeline karşı yorumlanır. Bayes faktörünün ilk kez bilimsel teorilerin karşılaştırılması kullanımını savunan Jeffreys'in (1961) kuralı, Bayes faktör yorumlamalarında yaygın olarak kullanılmaktadır. Buna göre: $1<B_{21}<3$ ise M2 modeli için olumlu fakat ihmal edilebilir bir kanıt bulunmaktadır. $3<B_{21}<10$ ise kanıt pozitiftir. $10<B_{21}<100$ ise kanıt kuvvetlidir. $B_{21}>100$ ise kanıt karar verici kuvvettedir.

Bundan sonraki kısımda model seçiminde Bayesian yaklaşımın pratik bir uygulaması olan Bayesian Information Criterian (BIC) genel olarak tanıtılacaktır.

Bayesian Information Criterian (BIC)

Model seçiminde Bayesian yaklaşımın kesin sonuçlar veren basit bir uygulaması BIC yakınsamasıdır (approximation). BIC'in pratikliği, onun standart istatiksel yazılım çıktılarından kolayca hesaplanabilmesinden dolayıdır. BIC'in diğer bir üstün yanı yuvalanmış modeller kadar yuvalanmış modellerin test edilmesine de uygun olmasıdır. Schwards (1978), BIC'i büyük örneklemlerde doğru modeli yüksek olasılıkla seçme prosüdürü sağlaması özelliği ile Bayesian yaklaşımın ötesinde bir geçerliğe sahip olduğunu savunmuştur.

Birden fazla modelin gözünüze alındğı durumlarda, her model ya bir null model, M0, ya da her verinin tamamen uyduğu 'doygun' (saturated) modelden, Ms, oluşan bir temel kriter (baseline) model ile ayrı ayrı karşılaştırılması gerekir. Temel kriter modelin doymuş model olması durumunda bu karşılaştırmayı aşağıdaki formül ile yapmak mümkündür.

$$BIC_k = DV_k - df_1 \log n \tag{8}$$

Bu formülde M1 modeli için BIC_k'nin değeri 2 $\log B_{sk}$ değerinde bir yakınsamadır. B_{sk} ise, M_k modeline karşı temel kriter model olan M_s modelidir. Formüldeki DV_k model M_k ile df_k arasındaki sapmanın (deviance) serbestlik derecesinin değeridir.

Eğer M_s modeli için BIC değeri sıfırdan küçükse, $BIC_k>0$, doymuş model M_k modeline tercih edilir. Eğer $BIC_k<0$ ise M_k modeli doymuş, M_s modeline karşılık tercih edilir. Bu durumda BIC_k değeri ne kadar (eksi işaretli olarak) küçükse M_k modelinin veriye uyumu o derece daha iyidir.

İki modelin karşılaştırıldığı durumlarda, örneğin M_j veM_k, aşağıdaki eşitlikler yazılabilir:

$$B_{sk}=p(D|M_j/p(D|M_k)$$

$$=\frac{[p(D|M_s)]}{[p(D|M_k)]} / \frac{[p(D|M_s)]}{[p(D|M_j)]}$$

$$=B_{sk}/B_{sj}, \text{ ve buna bağlı olarak}$$

$$2\log B_{jk}=2\log B_{sk}-2\log B_{sj}$$

$$\approx BIC_k-BIC_j \qquad (9)$$

Formül 9'dan görüldüğü üzere iki model BIC değerleri yoluyla karşılaştırabiliriz ve BIC değeri daha küçük olan diğerine karşılık kabul edilir.

Temel kriter modelin null hipotezi, M0, olması durumunda BICk işlemi BIC'k ile değiştirilir:

$$BIC'_k=-X^2{}_{ko}+p_k\log n \qquad (10)$$

Formüldeki X^2 ko, M0 modelinin Mk modeline karşı test eden likelihood-ratio test istatistiği; pk, test ile ilgili serbestlik derecesidir. Regrasyon tipi modellerde pk genellikle Mk modelindeki bağımsız değişken sayısının göstermektedir.

Temel kriter modelin null model olduğu durumda da formül 9'daki işlem geçerliliğini korur. Bu nedenle:

$$BICk-BICj=BIC'k-BIC'j \qquad (11)$$

yazmak mümkündür. Uygulamada BIC ya da BIC' kullanımı modelleri kestiren istatiksel yazılımların sapma mı yoksa LRT istatistiği mi hesapladığına göre değişir. Raftery (1994) her iki gösterge model seçimi ve test edilmesinde denk olduğunu ve fakat farklı anlamlar taşıdığını öne sürmektedir. BICk, modelin genel uyumunu ölçer. BIC'k ise, parametre sayısını değerlendirmeye yönelik olarak, Mk modelinin gözlenen sapmayı yeterince açıklayıp açıklayamadığı konusunda ölçüm bildirir.

SONUÇ

Bu çalışmada Bayesian yaklaşımın hipotez testi, model seçimi ve model belirsizliği konularındaki katkıları incelenmiştir. Model seçimine özel bir ağırlık verilmiştir. Ulaşılan bazı sonuçları bu şekilde sıralamak mümkündür.

-Bayesian yaklaşım, özellikle büyük örneklemlerde, hipotez testine yönelik P-değerlendiren daha kesin değerlendirme fırsatı sunmaktadır.

-Bayesian yaklaşım, yuvalanmamış doğrudan karşılaştırılması için daha basit bir yol sunmaktadır.

-Bayesian yaklaşım, çalışma konusu yapılan null ile ilgili kanıtı nicelleştirmektedir. Böylece null hipotezlerinin veri yetersizliğinden mi yoksa verinin null hipotezi yönünde kanıt sunduğu için mi red edilemediği konusunda açıklık getirmektedir.

-BIC ya da BIC'Bayes faktörüne basit ve doğru bir yakınsama sağlar.

-BIC sürekli model seçimine rehberlik eder.

-Bayesian yaklaşımın model seçimi ile ilgili olarak burada sunulan teknikleri standart istatistikleri çözümleme yazılımları ile kolayca elde edilebilir.

KAYNAKLAR

Berger, J.O. ve Sellke, T. (1987). "Testing a Point Null Hypothesis: the Irreconcialability of P Values and Evidence", *Journal of the American Statistical Associaton,* 82, pp. 112-122.

Bernardo, J.M. ve Smith, A.F.M. (1994). *Bayesian Theory,* New York: Wiley.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). *İscrete Multivaritate Analysis,* Cambridge, Mass.: MIT ress.

Fenech, A. ve Westfall, P. (1988). "The Power Function of onditional Log-linear Model Tests", Journal of the American Statistical Association, 83, pp. 198-203.

Freedman, D.A. (1983). "A Note on Screening Regression quations", *The American Statistician,* 37(2), pp. 152-155.

Freedman, D.A., Navidi, W.C.ve Peters, S.C. (1988). "On the Impact of Variable Selection in Fitting Regression Equations", *Model Uncertainty and its Statistical Implications,* (Ed. T.K. Dijkstra), Berlin: Springer-Verlag, pp. 1-16

Grusky, D.B. ve Hauser, R.M. (1984). "Comparative Social Mobility Revisitded: Models of Convergenge and Divergenge in 16 Countries", *American Sociologiucal Review,* 49,pp.19-38.

Hount, M. (1983). *Mobility Tables*, Beverly Hills, Claifornia: Sage Publications.

Hount, M. (1984). "Status, Autonomy and Tradining in Occupational Mobility", *American Journal of Sociology*, 89, 1379-1409.

Jeffreys, H. (1961). *Theory of Probability*, 3rd Edition, Oxford Universty Press.

Kass, R.E. ve Raftery, A.E. (1995). "Bayes Factors", *Journal of the American Statistical Association*,

Lee, P.M. (1989) Bayesian Statistics: An Introduction, Oxford Universty Press.

Miller, A.J. (1984). "Selection of Subsets of Regression Variables", *Journal of the Royal Statistical Society*, Series A, 147, pp. 389-425.

Miller, A.J. (1990). *Subset Selection in Regression*, New York: Champman-Hall.

Press, S.j. (1989) *Bayesian Statistics:: Principles, Models and Applications*, New York: Wiley.

Raftery, A.E.(1986). "Choosing Modeles for Cross-Classiffications", *American Sociological Review*,_51, pp. 145-146.

Raftery, A.E.(1983). *Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models*, Tecnical Report 255, Department of Statistics, Universty of Washington.

Schwarz, G.(1978). '"Estimating the Dimensions of a Model", *Annals of Statistics*, 6, pp. 461-464.