

Regresyon Çeşitleri

Probit ve Lojistik Regresyon

Tarımsal desteklerin ekim alanı üzerindeki etkisini ölçmeyi amaçladığımız bir çalışma varsayalım. Burada bağımlı değişken y ekim alanı, bağımsız değişken x ise tarımsal desteklerdir. Ekim alanı eğer üretici düzeyinde büyüklük olarak değişiyorsa, 10 dekar, 23 dekar, 243 dekar gibi, bu durumda diğer varsayımlar sağlandığında basit regresyon modeli kullanılabilir. Ancak bazı durumlarda bağımlı değişkenimiz buradaki örnek gibi sürekli bir değişken olmaz. Bağımlı değişkenlerin sürekli değişken yerine kategorik değişkenler olduğu durumlarda basit regresyon modeli yerine diğer alternatif hesaplama yöntemlerini kullanmak gerekir. Çünkü basit regresyon modeli bağımlı değişkende 0 ile 1 arasında değişen sınırlanmış bu yapıyı dikkate alamaz ve ek olarak normallik varsayımları sağlanamaz. Benzer örnekten ilerleyecek olursak tarımsal desteklerin ekim alanı büyüklüğü üzerine etkisi yerine bir ürünün yetiştirip yetiştirilmeme durumu üzerine etkisi ölçülebilir. Mısır desteğinin üreticilerin mısırı yetiştirip yetiştirmemesi üzerine etkisi gibi bir çalışmada bağımlı değişken mısırın yetiştirilmesi veya yetiştirilmemesi şeklinde ikili bir yapıya sahiptir. Üreticilerin mısır yetiştirdiği durumda bu değişken 1, yetiştirmediği durumda 0 değerini alır. İşte bu ikili yapıya sahip modellerin genel ismine doğrusal olasılık modelleri ve özelde ise sınırlı bağımlı değişken veya ikili çıktı modelleri denir. Probit ve lojistik regresyon da bu model grubu içerisinde yer alan iki regresyon çeşididir. Bu yöntemler olasılık hesaplamalarına dayanır ve bağımlı değişkenin 1 olması olasılığını bağımsız değişkenlerin fonksiyonu olarak hesaplar. Burada bağımsız değişken x 'in $y = 1$ olma olasılığı üzerindeki etkisi incelenir. Tarımsal destekler özelinde verilen desteklerin üreticinin mısır yetiştirme olasılığı üzerindeki etkisi hesaplanır. Probit ve lojistik modellerinde en çok olabilirlik adı verilen bir tahminci ile β parametreleri hesaplanır. Her iki modelde de hesaplanan parametre değerleri büyüklük ifade etmek amacıyla yorumlanmaz. Sadece bağımsız değişkenlere göre bağımlı değişkenin 1 olup olmama durumunun olasılığını ifade eder. Genelde araştırmacılar β katsayılarına dayanarak bağımsız değişkenlerin esneklik katsayısına benzeyen marjinal etkisini hesaplar ve bu değerler üzerinden yorumlarda bulunurlar. Probit ve lojistik regresyon hesaplamalarında marjinal etki sonuçları birbirlerine eşittir, bu yüzden iki model arasındaki seçim araştırmacıya kalmıştır.

Sansürlenmiş Regresyon

Sansürlenmiş regresyon veri setinin ve ilgilenilen değişkenin çeşitli nedenlere bağlı olarak kısıtlanması durumunda kullanılan bir yaklaşımdır. İnsanlara verilen maaşın onların çalışma süreleri üzerindeki etkisini ölçmeyi planlayan bir çalışmamız olsun. Veri setimizde eğer sadece çalışan insanlar varsa bu set çalışmayan insanları göz ardı eder. Bu durum araştırmacının araştırma sorusuna bağlı olarak belirli bir grup örneği seçtiğini ifade eder. Genel olarak bu tür yaklaşımlara örnek seçim modelleri ismi de verilir. Sansürleme veya verinin kısıtlanması konusu alttan veya üstten olabilir. Alttan ve üstten sansürlenmeye tarımsal örnekler verilebilir. Mısır ve pamuk üreticilerinde tarımsal desteklerin etkisini ölçmeyi amaçladığımız bir çalışmada her üretici her yıl hem mısır hem de pamuk ürününü yetiştirmediği için bu üreticilerin bağımlı değişkeni y birçok gözlem için sıfır değerini alacaktır. Bu durum verilerin alttan sansürlendiğini gösterir. Üstten sansürleme ise günümüzde uygulanan üretime bağlı fark ödemesi desteklerinde görülmektedir. Devlet çeşitli bölgeler için bir üst verim sınırı belirlemekte ve en fazla bu verime kadar destek vermektedir. ÇKS'den bir veri elde edildiğinde birçok üretici veriminin bu üst sınırdan olduğu görülecektir. Rakamsal olarak örnek vermek gerekirse, bir bölgede üst verim sınırı 1400 kilogram belirlendiğinde ve üretici 1500 kilogram yetiştirirse bu üreticinin desteklemeye tabi verim kaydı 1400 olarak kaydedilecektir. Her ne kadar üreticilerin belirli bir miktarı belirlenen üst verim sınırından daha yüksek üretim gerçekleştirebilirler de kayıtları bu üst sınır olan 1400'ü gösterecektir. Bu durum da verinin üstten sansürlendiğini ifade eder. Hem alttan hem de üstten sansürleme durumlarının araştırmacı tarafından dikkate alınması gerekmektedir. İşte sansürlenmiş regresyon da bu verinin özelliğini dikkate alınmakta ve ilgili model veri setinin bu özelliğine göre hesaplanmaktadır. Bu veri özelliğini dikkate almanın en temel yollarından birisi Heckman'ın seçim modeli veya sansürleme yaklaşımıdır. Mısır ve pamuk örneği ile devam edilecek olursa, bu yaklaşımda önce çeşitli bağımsız değişkenler ile üreticilerin mısır ve pamuk yetiştirip yetiştirmeme durumları probit model ile tahmin edilir. Burası yaklaşımın ilk aşamasıdır. Ardından hesaplanan probit sonuçlarına göre dağılım fonksiyonları hesaplanır ve bu dağılım fonksiyonları ters Mills oranı olarak tanımlanan bir değer hesaplanmasında kullanılır. Son ikinci aşamada ise model tahmin edilen örneğin desteğinin etkisini ölçtüğümüz bir denkleme bu hesaplanan ters Mills oranı da eklenir. Genel mantık üreticinin yetiştirip yetiştirmeme ihtimalinin yani sansürleme durumunun hesaplanıp bir düzeltme katsayısı olarak son modele eklenmesidir.

Görünürde İlişkisiz Regresyon

Basit regresyon modelinde birden fazla bağımsız değişken x 'in olduğu durumlara değinilmiş, ancak birden fazla bağımlı y değişkeni durumundan bahsedilmemiştir. Basit regresyon yaklaşımı örneğin mısır desteğinin mısır ekim alanı üzerindeki etkisini ölçmek için yeterli görünmektedir. Ancak araştırmacılar birden fazla ürün üzerinde etkiyi ürünler arası ilişkileri dikkate almak için ölçmek isteyebilirler. Çünkü üreticiler ekim alanı kararı verirken ve dolayısıyla tarım politikalarının etkisine maruz kalırken birden fazla ürün arasından tercih yaparlar. Bu durumda ürünler arası tamamlayıcı veya rekabetçi ilişkilerin dikkate alınması gerekebilir. Benzer örnekle devam edilirse, mısır

Regresyon Çeşitleri

desteğinin mısır ekim alanı yanı sıra pamuk ekim alanı üzerindeki etkisi nedir, aynı şekilde pamuk desteğinin mısır ekim alanı üzerindeki etkisi nedir ve mısır ile pamuk ürünlerinin rekabetçi ürünler olması hesaplamalarda nasıl dikkate alınabilir şeklinde araştırma soruları oluşturulabilir. Bu örnekte mısır ve pamuk ekim alanı olmak üzere iki tane y dolayısıyla iki tane bağımlı değişken vardır. Bu durumda araştırmacılar şu şekilde denklem serileri oluşturur:

$$y_1 = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + u_1$$

$$y_2 = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + u_2$$

Burada y_1 mısır ekim alanını, y_2 pamuk ekim alanını, x_1 mısır desteğini, x_2 ise pamuk desteğini ifade etmektedir. β_{11} mısır desteğinin mısır ekim alanı, β_{12} pamuk desteğinin mısır ekim alanı, β_{21} mısır desteğinin pamuk ekim alanı, β_{22} pamuk desteğinin pamuk ekim alanı üzerine etkisini gösterir. Ürünler arası rekabetçi ilişki nedeniyle u_1 ve u_2 hata terimlerinin de birbirinden bağımsız olmadığı varsayılır.

Bu iki denklem ayrı ayrı en küçük kareler yöntemi kullanılarak hesaplanabilir. Ancak bu yaklaşım iki ürün arasındaki ilişkiyi göz ardı eder. İki ürünün rekabetçi olduğunu varsaydığımızda bu denklemlerin hata terimlerinin ilişkili olduğu ifade edilir. Bu durumda ayrı ayrı ölçmek yerine iki denklemi aynı anda ölçmek ve rekabetçi ilişkiyi dikkate almak daha uygun bir yaklaşım olur. Bu ölçümün gerçekleştirilmesi için kısaca β katsayılarının hesaplanmasında genelleştirilmiş en küçük kareler yöntemi kullanılır. Bu yaklaşımda her bir denklem katsayıları ayrı ayrı en küçük kareler yöntemi ile tahmin edilir. Tahmin sonrasında hata terimleri hesaplanır. Ardından hata terimlerinin kovaryans matrisleri oluşturulur ve bu matris en sonda bütün denklem serisinin tahmincisini hesaplayacak formüle eklenir:

$$\hat{\beta} = [x'\Omega^{-1}x]^{-1}x'\Omega^{-1}y$$

Burada $\Omega = \Sigma \otimes I$ 'dir. Σ hata terimleri kovaryans matrisini, I gözlem sayısı büyüklüğündeki birim matrisi, \otimes kronocker çarpımını ifade eder. Görünürde ilişkisiz regresyon tarımın doğasında bulunan ürünler arası ilişkileri dikkate aldığı için çeşitli çalışmalarda tarımsal desteklerin analizi kapsamında kullanılmıştır.

Panel Veri Model ve Birey Etki

Veri bölümünde bahsedildiği gibi günümüzde artan veri toplama imkânlarına bağlı olarak panel veriler araştırmalarda sıklıkla kullanılan veriler olmuştur. Tarım açısından bu veri de aynı çiftçileri zaman içerisinde izler ve çiftçilerin o yılı ve devam eden yıllardaki verilerini toplarız. Seçilen bir bölgede mısır üreticilerinin 2008-2017 yılları arası mısır üretim ve destek verileri panel veriye örnek verilebilir. Her üreticiyi her yıl ilgili değişken açısından bulabilir isek dengeli bir panel veri elde eder, aksi durumda üreticilerin çeşitli yıllarda veri setinden çıkması durumunda dengesiz bir panel veri elde ederiz. Panel veri modellerinin hesaplamalarını yapan bilgisayar paket programları dengeli ve dengesiz panelleri sorunsuz bir şekilde hesaplayabilmektedir. Araştırmacı açısından dikkat edilmesi gereken nokta veri setinin dengesiz olması durumunun tesadüfi olarak gerçekleşmesidir. Veri setinden ayrılan veya çıkan üreticiler araştırma sorusunu etkileyecek bir nedene bağlı olmamalıdır. Panel veri ve hesaplama yöntemlerinin temel noktası sabit etki adı verilen bir konudur. Üreticilerin zaman içerisinde izlenebilme olanağı onların sahip oldukları ve araştırmacı tarafından gözlenemeyen etkilerin dikkate alınabilmesine yardımcı olur. Örneğin üreticilerin mısır yetiştirmedeki yeteneği araştırmacı tarafından gözlenemeyebilir. Ancak panel veri yöntemleri ile bu yetenek gibi gözlenemeyen konular hesaplamalarda kontrol altına alınabilir. Böylelikle araştırmacı ilgilendiği değişkenin etkisi konusunda daha kapsamlı ve net çıkarımlarda bulunabilir. Panel veri modellerinde ve ilgili hesaplama yöntemlerinin seçiminde dikkat edilmesi gereken konu bahsedilen bu sabit etkinin bağımsız değişkenler ile ilişkili olup olmamasıdır. Bir bölgede mısır yetiştiren üretici yetenekli olduğu için diğer üreticilerden fazla üretim yapabilir ve daha fazla destek alabilir. Desteğin etkisini ölçen bir araştırmacı bu yeteneği dikkate almaz veya kontrol etmez ise tahmin edeceği destek değişkeninin etkisi gerçek politika etkisinden farklı olabilir. Panel veri yöntemlerinde sabit etkiler burada bahsedildiği gibi bağımsız değişkenler ile ilişkili olursa sabit etkiler modeli denilen bir hesaplama yöntemi ile bu sabit etkiler konusu dikkate alınır. Eğer bireylerin sabit etkileri örneğin burada bahsedildiği gibi yeteneği, açıklayıcı değişkenler ile yani destekler ile ilişkisiz ise o zaman tesadüfi etkiler modeli adı verilen bir yöntem ve hesaplama tekniği kullanılır. Her ne kadar farklı yaklaşım teknikleri bulunsada, çoğu durumda bağımsız değişkenler için hesaplanan standart hataların düzeltilmesi gerekir.