

Davranış Bilimlerinde İstatistik

Regresyon Analizi

Doç. Dr. Seher YALÇIN

Regresyon Analizi (Köklü ve diğ., 2006)

- Regresyon, bir deęişkene ilişkin ölçümlerin grup ortalamasına doğru çekilmesidir.
- Regresyon analizi, aralarında ilişki olan iki ya da daha fazla deęişkenden birinin bağımlı deęişken, diğerlerinin bağımsız deęişkenler olarak ayrımı ile aralarındaki ilişkinin matematiksel bir eşitlik ile açıklanması sürecidir.

Regresyon Analizi (Köklü ve diğ., 2006)

- Bağımsız/Yordayıcı Değişken: Genellikle x ile gösterilir. Başka bir değişken tarafından etkilenmeyen ama y 'nin nedeni olan ya da onu etkilediği düşünülen (açıklayıcı) değişkendir.
- Bağımlı/Yordanan Değişken: Genellikle y ile gösterilir. x değişkenine bağlı olarak değişebilen ya da ondan etkilenen (açıklanan) değişkendir.

Regresyon Analizi (Köklü ve diğ., 2006)

- Bağımlı/Yordanan değişken, bir değişken tarafından tahmin ediliyorsa **basit regresyon**, iki ya da daha fazla değişken tarafından tahmin ediliyorsa **çoklu regresyon** denir.
- Eğer bağımlı değişken birden fazla ise **çok değişkenli regresyon** denir.

Araştırma Soruları

- Yaş değişkeni okuma başarısının manidar bir yordayıcısı mıdır? (*Basit Regresyon*)
- Yaş, anne eğitim durumu, sosyo-ekonomik düzey (sed) okuma başarısının manidar bir yordayıcısı mıdır? (*Çoklu Regresyon*)

Basit Regresyon için denklem;

- $Y = a + bX_i + e_i$
- Y: bağımlı değişken
- X: bağımsız değişken
- b: regresyon katsayısı (regresyon doğrusunun eğimi, x'deki bir birim değişiminin Y'de yol açtığı ortalama değişim miktarıdır)
- a: kesim noktasını gösteren sabit
- e_i : hata

Çoklu Doğrusal Regresyon denklemi;

- $Y = aX_1 + bX_2 + cX_3 + \dots + k$
- Y: bağımlı değişken
- X: bağımsız değişken
- a, b, c : değişkenin katsayısı
- k: sabit

Çoklu Doğrusal Regresyon Analizi

- Çoklu regresyon analizinde hem elde edilen regresyon denklemi yardımıyla tahminde bulunmak, hem de değişkenlerden hangilerinin modele daha fazla katkı yaptığını belirlemek amaçlanır (Alpar, 2011).
- Standartlaştırılmış regresyon katsayılarına **beta katsayıları (β)** da denir. Beta katsayısı büyük olan değişken, modele en fazla katkıyı yapan değişkendir (Kline, 2005).

Çoklu Doğrusal Regresyon Analizi

- β^2 değeri ise, bir X_i değişkeninin Y 'de diğer değişkenler tarafından paylaşılamayan varyans miktarıdır.
- Bağımsız değişkenlerin bağımlı değişken üzerindeki ortak etkisi \mathbf{R}^2 ile incelenir. Değişkenlerin birlikte Y 'de açıkladıkları varyans oranını verir.

Çoklu Doğrusal Regresyon Analizinde Değişken Ekleme/Seçme Yöntemleri

- **Enter yöntemi**
- Değişkenlerin tümü modele dâhil edilerek modelde bağımlı değişkenin yordanma düzeyi test edilmektedir. Modele önemli katkısı olmayan değişkenler modelden çıkarılır ve en iyi regresyon denklemi bulunana kadar işleme devam edilir (Tabachnick ve Fidell, 1996).
- **İleriye Doğru Seçim/Değişken Ekleme Seçimi (Forward Selection)**
Değişken seçme işlemine sadece sabit terimin bulunduğu bir model ile başlanır. Ardından en yüksek ilişkili değişkenden başlanarak modele tek tek eklenir.

Çoklu Doğrusal Regresyon Analizinde Değişken Ekleme/Seçme Yöntemleri

- **Geriye Doğru Çıkarma Yöntemi (Backward Selection)**
- Tüm bağımsız/yordayıcı değişkenler modele dahil edilerek başlanır ve «manidar olmayan" yordayıcı değişkenler teker teker modelden çıkarılır.

Çoklu Doğrusal Regresyon Analizinde Değişken Ekleme/Seçme Yöntemleri

- **Adım Adım/Aşamalı Regresyon Yöntemi (Stepwise Selection)**
- Bağımlı/yordanan değişken ile en yüksek korelasyonu veren bağımsız değişken ilk olarak analize alınmaktadır. Diğer bağımsız/yordayıcı değişkenler, varyansı açıklama oranına göre en yüksekte başlanarak sırasıyla modele eklenmektedir.
- Bağımlı değişkeni yordamada manidar katkıda bulunan bağımsız değişkenler sadece denkleme alınır (Tabachnick ve Fidell, 1996).

Çoklu Doğrusal Regresyon Analizinde Değişken Ekleme/Seçme Yöntemleri

- Çoklu bağlantı olduğu durumlarda *geriye doğru seçim* yöntemini, çoklu bağlantı olmadığı durumlarda ise *adım-adım regresyon yönteminin* uygulanması genelde kabul görmektedir.

Çoklu Doğrusal Regresyon Analizi Varsayımlar

(Büyüköztürk, 2011; Kalaycı, 2010)

- Değişkenler en az eşit aralık ölçeğinde olmalı
- Bağımlı değişkenler sürekli olmalı
 - Sürekli olmayan bağımsız değişkenler “dummy” değişkeni olarak tanımlandıktan sonra analize alınabilir.
- Bağımsız değişkenler arasındaki korelasyon katsayısı yüksek olmamalı
- Çok değişkenli normal dağılım varsayımı sağlanmalı,

Çoklu Doğrusal Regresyon Analizi Varsayımlar (devam) (Büyüköztürk, 2004; Kalaycı, 2010)

- Doğrusallık varsayımı sağlanmalı,
- Yordayıcı değişkenlerle bağımlı değişken arasında doğrusal bir ilişki olmalı,
- Hata terimleri arasında ilişki olmaması (otokorelasyon) ve hata değerlerinin normal dağılması gereklidir.

Çoklu Bağlantılılık (Multi-collinearity) (Büyüköztürk, 2004)

- Çoklu bağlantılılık, yordayıcı değişkenler arasındaki ilişkinin yüksek olmasıdır.
- 1. Korelasyon analizi sonucunda elde edilen yüksek korelasyonlar (.90 üzeri).
- 2. Bir bağımsız değişkenle ilgili olarak diğer bağımsız değişkenlerin açıklayamadıkları varyans oranı olan tolerans değerinin $(1 - R^2)$ değerinin .20'den düşük olması

Çoklu Bağlantılılık (Multi-collinearity) (Büyüköztürk, 2004)

- 3. Varyans artış faktörü (VIF = Variance Inflation Factors) değerinin 10'un altında olması istenen koşuldur. Ancak 10-30 arasında yer alan VIF değerleri de kabul görmektedir.
- 4. Durum indeks (condition indices, CI) değeri: en yüksek öz değer in değişkenin öz değerine oranının kareköküdür, CI değerinin 30'dan yüksek çıkması durumunda çoklu bağlantı sorunu vardır.

Çoklu Bağlantılık (Multi-collinearity) olması durumunda arařtırmacılar (Büyüköztürk, 2004);

- Yeniden veri toplayabilir ve analizi tekrarlayabilir.
- Temel Bileşenler Analizi yaparak bu deęişkenlerden yeni bir deęişken üretebilir ve analize bu deęişken alınır.
- VIF deęeri en yüksek toleransı en düşük bağımsız deęişken modelden çıkartılarak analiz tekrar edilebilir.
- Ridge regresyon yöntemi denenebilir.

Otokorelasyon

- Otokorelasyon, hata terimleri arasında ilişki olmamasıdır.
- Herhangi bir değişken için hesaplanıp, bu değişkenin gözlem değerlerinin kendinden önceki ve sonraki gözlem değerleri ile ilişkisi olup olmadığını varsa derecesini bildirir.
- Bu katsayının, pozitif bir değer olarak 1'e yaklaşması pozitif otokorelasyon, -1'e yaklaşması ise negatif otokorelasyonun bulunduğu göstergesidir.

Otokorelasyon

- **Otokorelasyon Nasıl test edilir?**
- Durbin-Watson Testi (0-4 arasında, 1'den küçük ve 3'den büyük değerler sorun var)
- **Nedenleri**
- Dahil Edilmeyen Değişkenler (önemli bazı değişkenlerin ihmali)
- Ölçme Hataları

Kaynaklar

- Alpar, R. (2011). *Spor, sađlık ve eđitim bilimlerinde uygulamalı istatistik ve geđerlik-güvenirlik*. Ankara: Detay Yayıncılık.
- Büyüköztürk, Ş. (2004). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem A Yayıncılık.
- Kalaycı, Ş. (2010). *SPSS Uygulamalı Çok Deđişkenli İstatistik Teknikleri (5. Baskı)*. Ankara: Asil Yayın Dađıtım.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling (2nd ed.)*. New York: Guilford.
- Köklü, N., Büyüköztürk, Ş. & Çokluk Bökeođlu, Ö. (2006). *Sosyal Bilimler için İstatistik*. Ankara: Pegem Yayınları.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics (3rd ed.)*. New York Harper Collins.