


Bağlanım Çözümlemesi

Yrd. Doç. Dr. A. Talha YALTA
Ekonometri 1 Ders Notları
Sürüm 2,0 (Ekim 2011)



Açık Lisans Bilgisi

İşbu belge, “Creative Commons Attribution-Non-Commercial ShareAlike 3.0 Unported” (CC BY-NC-SA 3.0) lisansı altında bir açık ders malzemesi olarak genel kullanıma sunulmuştur. Eserin ilk sahibinin belirtilmesi ve geçerli lisansın korunması koşulu ile özgürce kullanılabilir, çoğaltılabilir ve değiştirilebilir. Creative Commons örgütü ve “CC-BY-NC-SA” lisansı ile ilgili ayrıntılı bilgi “<http://creativecommons.org>” adresinde bulunmaktadır. Ekonometri ders notlarımın güncel sürümüne “<http://yalta.etu.edu.tr>” adresinden ulaşabilirsiniz.

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011 

Ders Planı

- 1 **Temel Kavramlar**
 - Baęlanım Teriminin Anlamı
 - Ekonometrik Çözümlemede Kullanılan Verilerin Nitelięi
- 2 **Varsayımsal Bir Örnek**
 - Koşullu Olasılık ve Koşullu Ortalama
 - Anakütle Baęlanım İşlevi
 - Örneklem Baęlanım İşlevi

Ders Planı

1 Temel Kavramlar

- Bağlanım Teriminin Anlamı
- Ekonometrik Çözümlemede Kullanılan Verilerin Niteliği

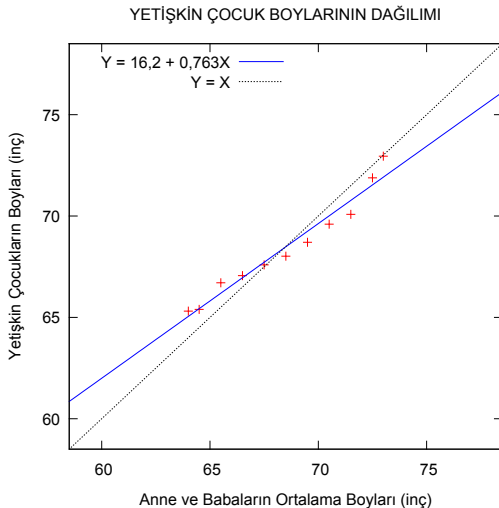
2 Varsayımsal Bir Örnek

- Koşullu Olasılık ve Koşullu Ortalama
- Anakütle Bağlanım İşlevi
- Örneklem Bağlanım İşlevi

Bağlanım Teriminin Anlamı

- İngilizce “regression” teriminin sözcük anlamı, istatistikteki “**sıradanlığa doğru çekilme**” (regression toward mediocrity) olgusundan gelmektedir.
- Bu terim ilk kez İngiliz antropolog, meteorolojist, kaşif, mucit ve istatistikçi Sir Francis Galton (1822 - 1911) tarafından kullanılmıştır.
- Galton ünlü bir yazısında belli bir boydaki anne-babaların yetişkin çocuklarının ortalama boylarının genel nüfustaki ortalama boya çekilme eğiliminde olduğunu bulmuştur.
- Günümüzde kullanılan anlamıyla “regression” bağımlı bir değişkeni, tahmin ya da çıkarım amacıyla farklı bağımsız değişkenler ile ilişkilendiren istatistiksel bir yöntemdir.
- Bu terimin uygun ve doğru Türkçe karşılığı ise “**bağlanım**” sözcüğüdür (Bkz. TDK İstatistik Terimleri Sözlüğü).

Sıradanlığa Çekilme Kavramı



Bağlanım Sözcüğünün İstatistikteki Yorumu

Bağlanım terimi istatistikte bir çözümleme yöntemini anlatır:

Bağlanım Çözümlemesi

Bağlanım çözümlemesi, bir bağımlı değişkenin başka açıklayıcı değişkenlerle olan ilişkisini, birincinin ortalama değerini ikinci(ler)in bilinen ya da sabit değerleri cinsinden tahmin etme ya da kestirme amacıyla inceleyen bir istatistiksel yöntemdir.

- Diğer bir deyişle, bağlanım yöntemi, bağımlı değişkendeki değişiklikleri açıklayıcı değişken denilen çeşitli etmenleri denetim altında tutarak inceler.
- Bağlanım çözümlemesindeki ilgi odağı kesin ilişkiler değil istatistiksel ilişkilerdir.
- Kullanılan değişkenler genellikle rastsal ya da “olasılıksal” (stochastic) ya da olasılık dağılımı olan değişkenlerdir.

Bağlanım ile İlgili Temel Terimler

- Bağlanım çözümlemesinde kullanılan sol ve sağ yan değişkenleri yazında farklı adlar ile karşımıza çıkabilirler:

SOL YAN (Y)		SAĞ YAN (X)	
Türkçe	İngilizce	Türkçe	İngilizce
"Açıklanan değişken"	(Explained variable)	"Açıklayıcı değişken"	(Explanatory variable)
"Bağımlı değişken"	(Dependent variable)	"Bağımsız değişken"	(Independent variable)
"Bağlanan"	(Regressand)	"Bağlayan"	(Regressor)
"Kestirilen"	(Predictand)	"Kestiren"	(Predictor)
"Tepki değişkeni"	(Response variable)	"Denetim değişkeni"	(Control variable)
"İçsel değişken"	(Endogenous variable)	"Dışsal değişken"	(Exogenous variable)

Bağlanım ve Nedensellik

İstatistiksel bir ilişki kendi başına bir nedensellik anlamı taşımaz. M. G. Kendal ve A. Stuart'ın sözleriyle:

“İstatistiksel bir ilişki ne denli güçlü ve ne denli anlamlı olursa olsun, asla nedensel bir ilişki kuramaz. Bizim nedensellik düşüncelerimiz istatistiğin dışından, eninde sonunda şu ya da bu kuramdan gelmelidir.”

Bağlanım ve İlinti

- “İlinti” (correlation) çözümlemesi, iki değişken arasındaki doğrusal ilişkinin gücünü inceler.
- Bağlanım çözümlemesi ve ilinti çözümlemesi yakından ilişkili olsa da bu iki yöntem arasında önemli kavramsal farklar vardır.
- İlinti çözümlemesinde herhangi iki değişken “bakışimli” (symmetric) olarak ele alınabilir.
- Diğer bir deyişle bağımlı ve açıklayıcı değişkenlerden söz edilmez.
- Bağlanım çözümlemesinde ise değişkenlerin ele alınışı tek yönlüdür. Bağımlı değişkenin olasılıksal olduğu, açıklayıcı değişken(ler)in ise değişmeyen değerler aldığı varsayılır.

Veri Seti Türleri

“Görgül” (empirical) çözümlemelerde üç tür veri seti kullanılır:

- 1 “Zaman serisi” (time series) veri setleri
- 2 “Yatay-kesit” (cross-sectional) veri setleri
- 3 “Karma” (pooled) veri setleri

Zaman Serileri

Zaman Serileri

Zaman serisi, bir değişkenin farklı zamanlarda gözlenen bir değerler setidir.

Zaman serilerine örnek olarak aşağıdakiler gösterilebilir:

- Hisse senedi fiyatları (günlük / dakikalık)
- Para arzı (haftalık)
- Tüketici Fiyat Endeksi (aylık)
- Gayri Safi Milli Hasıla (üç aylık)
- Hükümet bütçesi (yıllık)
- Genel seçim sonuçları (dört yıllık)

Yatay-Kesit Verileri

Yatay-Kesit Verileri

Yatay-kesitsel veriler, zaman içinde belli bir noktada derlenerek oluşturulan veri setleridir.

Yatay-kesit verilerine örnek olarak şunlar gösterilebilir:

- TÜİK tarafından belli aralıklarla düzenlenen tüketici harcamaları anketi
- Çeşitli kurumlarca yürütülen kamuoyu araştırmaları
- Hisse senedi fiyatlarının belli bir gün sonundaki değerleri

Karma Veriler ve Panel Verileri

Karma Veriler

Karma veriler, hem zaman serisi hem de yatay-kesit ögeleri içeren verilerdir.

- Karma verilere örnek olarak çeşitli illere ait gelir, işsizlik, iç göç gibi istatistikleri içeren bir veri seti gösterilebilir.

“Panel” (panel) verileri denen özel bir karma veri tipi vardır:

Panel Verileri

Birden fazla değişkenin zaman içerisinde izlenilmesi ile ortaya çıkan veri seti türüdür.

- Panel verilerine örnek olarak ABD Michigan Üniversitesi tarafından düzenlenen Panel Study of Income Dynamics (PSID) veri tabanı gösterilebilir.

Aldıkları Değerlere Göre Veri Türleri

Aldıkları değerler bakımından ise veriler ikiye ayrılırlar:

Nicel Veriler

Gelir, fiyatlar, para arzı, faiz oranları ...

Nitel Veriler

Erkek / kadın, evli / bekar, üniversite mezunu / değil, ...

Verilerin Doğruluk Derecesi

Ekonomik araştırmalarda kullanılan veriler çoğu zaman nitelik yönünden çok iyi düzeyde olamayabilmektedirler:

- Çoğu toplum bilim verileri deneysel olmadığı için gözlem hataları içermektedir.
- Deneysel verilerde bile ölçüm hataları olabilmektedir.
- Anketle toplanan verilerde yanıt alamama sorunu ya da **“seçim yanlılığı”** (selection bias) doğabilmektedir.
- Kullanılan örnekleme yöntemi **“örnekleme yanlılığı”** (sampling bias) sorununa yol açabilmektedir.
- **“Toplulaştırmalı”** (aggregated) iktisadi veriler hane halkı gibi mikro birimler için fazla açıklayıcı olamayabilmektedir.

Sonuç olarak; ekonometrik yöntemlerin başarısı kullanılan verilerin kaynak, nitelik ve doğruluk derecesine bağlıdır.

Ders Planı

- 1 Temel Kavramlar
 - Bağlanım Teriminin Anlamı
 - Ekonometrik Çözümlemede Kullanılan Verilerin Niteliği
- 2 Varsayımsal Bir Örnek
 - Koşullu Olasılık ve Koşullu Ortalama
 - Anakütle Bağlanım İşlevi
 - Örneklem Bağlanım İşlevi

Varsayımsal Bir Örnek

- Bağlanım çözümlemesine başlangıç olarak ikili bağlanım modelini inceleyeceğiz.
- İki değişkenli durum çoğu uygulama için yetersiz olsa da temel bilgileri olabildiğince yalın gösterebilmek açısından önemlidir.
- İkili bağlanıma varsayımsal bir örnek olarak toplam nüfusu 60 aileden oluşan bir ülke düşünelim.
- Bu ailelerin vergiden sonraki harcanabilir haftalık gelirleri X ve haftalık tüketim harcamaları Y arasındaki ilişkiyi tahmin etmek istiyor olalım.
- Bunun için öncelikle bu 60 aileyi gelirleri yaklaşık aynı olan 10 farklı öbeğe ayıralım.

Varsayımsal Örnek Verileri

- Örneğimiz ile ilgili varsayımsal veriler aşağıdadır:

Çizelge: Haftalık Aile Geliri X ile Haftalık Tüketim Harcamaları Y , \$

$Y \downarrow, X \rightarrow$	80	100	120	140	160	180	200	220	240	260
	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
Toplam	325	462	445	707	678	750	685	1043	966	1211

- Buradaki her bir sütun, farklı gelir düzeylerine (X) karşılık gelen tüketim harcamaları (Y) dağılımını göstermektedir.

Koşullu Olasılık ve Koşullu Ortalama

- Örnekteki $X = 80$ değerine karşılık gelen 5 ayrı Y değeri bulunmaktadır: 55, 60, 65, 70 ve 75.
- Yukarıdaki tüketim harcamalarının her birinin gerçekleşme olasılığı ise $\frac{1}{5}$ 'tir.
- Bu durumda, $X = 80$ olduğunda Y 'nin de 55 olma **“koşullu olasılığı”** (conditional probability) $P(Y = 55 | X = 80) = \frac{1}{5}$ 'tir.
- **“Koşullu ortalama”** (conditional mean) ya da **“koşullu beklenen değer”** (conditional expected value) ise Y 'nin her bir koşullu olasılık dağılımı için beklenen değerini gösterir.
- Koşullu ortalamayı bulmak için ilgili Y değerleri ve bunlara karşılık gelen koşullu olasılıklar çarpılıp toplanır.
- Örnek olarak, $X = 80$ iken Y 'nin koşullu ortalaması $55(\frac{1}{5}) + 60(\frac{1}{5}) + 65(\frac{1}{5}) + 70(\frac{1}{5}) + 75(\frac{1}{5}) = 65$ olur.

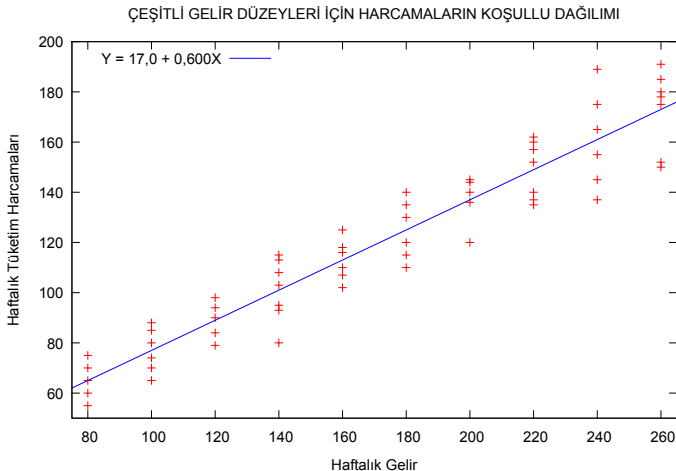
Koşullu Olasılık ve Koşullu Ortalama

Çizelge: $P(Y|X_i)$ Koşullu Olasılık ve Koşullu Ortalamaları

$Y \downarrow, X \rightarrow$	80	100	120	140	160	180	200	220	240	260
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
	–	1/6	–	1/7	1/6	1/6	–	1/7	1/6	1/7
	–	–	–	1/7	–	–	–	1/7	–	1/7
Ortalama	65	77	89	101	113	125	137	149	161	173

Anakütle Bağlanım İşlevi

Verilerimizi “serpılım çizimi” (scatter plot) üzerinde inceleyelim:



Anakütle Bağlanım İşlevi

Çizimde görülen artı eğimli doğrunun gösterdiği matematiksel işlev “**anakütle bağlanım işlevi**” (population regression function) ya da kısaca “**ABI**” (PRF) olarak adlandırılır:

Anakütle Bağlanım İşlevi

Anakütle bağlanım işlevi, açıklayıcı değişken(ler)in sabit değerlerine karşılık gelen bağımlı değişkenin koşullu ortalamaları ya da koşullu beklenen değerlerinin geometrik yerini gösterir.

- Her koşullu ortalama X 'in bir işlevidir: $E(Y|X_i) = f(X_i)$.
- Anakütle bağlanım işlevi denilen $f(X_i)$, X 'teki değişmeye karşılık Y 'nin dağılımının ortalama tepkisini vermektedir.
- Kısaca Y 'nin ortalama değeri ile ilgileniyoruz ama özellikle de Y 'nin ortalamasının X 'lere bağlı olarak nasıl değiştiğini bulmaya çalışıyoruz.

Anakütle Bağlanım İşlevi

- $f(X_i)$ 'nin işlev biçiminin ne olduğu sorusu önemlidir.
- Gerçek yaşamda tüm anakütle incelemeye açık olmadığı için burada iktisat kuramından yararlanılmalıdır.
- Örnek olarak, bir ekonomist tüketim harcamalarının gelire doğrusal bir ilişki içinde olduğunu söylüyor olsun.
- Bu durumda varsayılabilecek doğrusal işlev de şu olur:

$$E(Y|X_i) = f(X_i) = \beta_1 + \beta_2 X_i$$

Doğrusal İşlevin Anlamı

“Doğrusal” (linear) işlev, “değişkenlerde doğrusallık” (linearity in the variables) ve “değiştirgelerde doğrusallık” (linearity in the parameters) olmak üzere iki farklı anlama gelebilir:

Değişkenlerde Doğrusallık

Doğal ve basitçe bağlanım işlevinin düz bir doğruyu gösterdiği durumdur.

Doğrusal: $Y_i = \beta_1 + \beta_2 X_i$

Doğrusal-dışı: $Y_i = \beta_1 + \beta_2 X_i^2$

Değiştirgelerde Doğrusallık

$E(Y|X_i)$ 'nin β değiştirgelerinin doğrusal bir işlevi olduğu durumdur.

Doğrusal: $Y_i = \beta_1 + \beta_2 X_i^2$

Doğrusal-dışı: $Y_i = \beta_1 + \sqrt{\beta_2} X_i$

Rastsal Hata Terimi

- Örneğimizde görüldüğü gibi gelir artarken tüketim harcamaları da genel olarak artmaktadır.
- Diğer yandan, tekil bir ailenin harcamasının geliri daha düşük olan bir aileden fazla olması da zorunlu değildir:

Çizelge: Haftalık Aile Geliri X ile Haftalık Tüketim Harcamaları Y , \$

$Y \downarrow, X \rightarrow$	80	100	120	140	160	180	200	220	240	260
	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
Toplam	325	462	445	707	678	750	685	1043	966	1211

Rastsal Hata Terimi

- Tekil bir ailenin harcamasının aynı gelir düzeyindeki bütün ailelerin harcamalarının ortalaması, diğer bir deyişle koşullu beklenen değeri dolayında dağıldığını biliyoruz.
- Buna göre, bireysel Y_i 'nin kendi beklenen değerinden gösterdiği “sapma” (deviation) şöyle gösterilebilir:

$$u_i = Y_i - E(Y|X_i)$$

ya da $Y_i = E(Y|X_i) + u_i$

ya da $Y_i = \beta_1 + \beta_2 X_i + u_i$

- Buradaki u_i “bozukluk” (disturbance) terimi, artı ya da eksi değerler alabilen ama gözlenemeyen “rastsal hata terimi” (random error term) diye adlandırılır.

Rastsal Hata Teriminin Beklenen Değeri

- $Y_i = E(Y|X_i) + u_i$ eşitliğinin her iki yanının beklenen değeri alınırsa şu bulunur:

$$\begin{aligned}Y_i &= E(Y|X_i) + u_i \\E(Y_i|X_i) &= E[E(Y|X_i)] + E(u_i|X_i) \\E(Y_i|X_i) &= E(Y|X_i) + E(u_i|X_i) \\0 &= E(u_i|X_i)\end{aligned}$$

- $E(Y_i|X_i)$ ile $E(Y|X_i)$ aynı şey olduğu için, $E(u_i|X_i) = 0$ olur.
- Bu durumda, u_i 'lerin koşullu ortalamasının sıfır olduğu varsayımına dayanılarak, bağlanım doğrusunun Y 'nin koşullu ortalamasından geçtiği sonucuna ulaşılabilir.

Rastsal Hata Teriminin Önemi

Modele katılmayan ama Y 'yi etkileyen tüm değişkenlerin yerine geçen hata terimi u_i 'nin modele açıkça koyulması gereğinin nedenlerinden bazıları şunlardır:

- 1 Kuramın belirsizliği ya da eksikliği
- 2 Yeterli ya da geçerli verilerin bulunamaması
- 3 İlişkili ancak ortak etkisi küçük olan değişkenler
- 4 İnsan davranışlarının doğasında olan rastsallık
- 5 Güçsüz “yaklaşık değişkenler” (proxy variables)
- 6 Basitlik ilkesi
- 7 Bilinemeyen işlev biçimi

Örneklem Bağlanım İşlevi

- Gerçek yaşamda anakütle verilerine ulaşabilme olasılığı düşüktür.
- Çoğu uygulamada elimizde yalnızca anakütleden alınmış örneklem verileri bulunmaktadır.
- Öyleyse yanıtlamamız gereken önemli soru, örneklem verilerini kullanarak anakütle bağlanım işlevi ABİ'yi tahmin edip edemeyeceğimiz sorusudur.
- Rastsal bir örneklem kullanarak bulunan bağlanım işlevine “**örneklem bağlanım işlevi**” (sample regression function) ya da kısaca “**ÖBİ**” (SRF) denir.
- Bu işlevi anlatan doğruya ise “**örneklem bağlanım doğrusu**” (sample regression line) adı verilir.

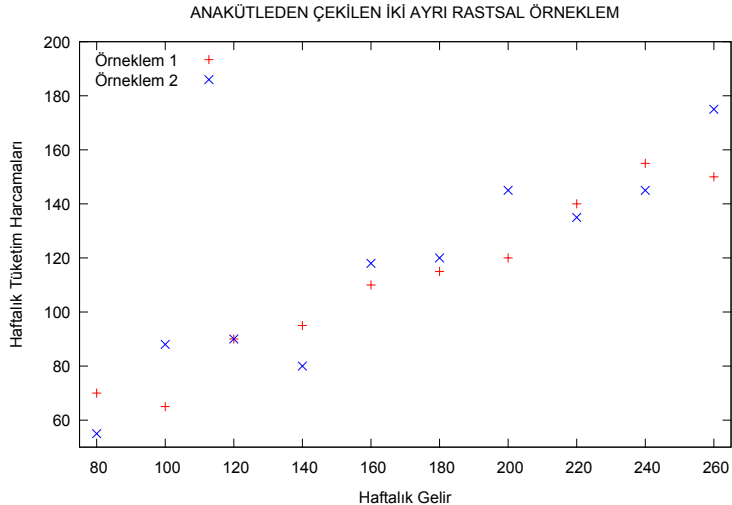
Örneklem Bağlanım İşlevi

Anakütleden her biri 10 gözlem büyüklüğünde iki farklı rastsal örneklem çekelim:

Çizelge: Anakütleden Çekilmiş İki Rastsal Örneklem

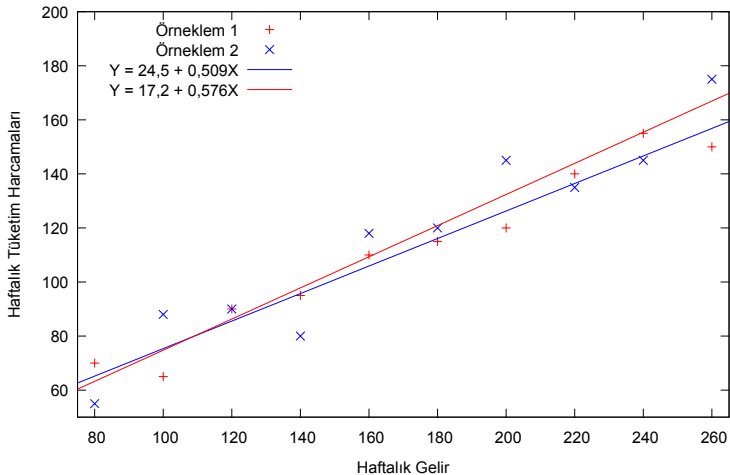
X	Y	X	Y
80	70	80	55
100	65	100	88
120	90	120	90
140	95	140	80
160	110	160	118
180	115	180	120
200	120	200	145
220	140	220	135
240	155	240	145
260	150	260	175

Örneklem Bağlanım İşlevi



Örneklem Bağlanım İşlevi

İKİ AYRI ÖRNEKLEME DAYANAN İKİ FARKLI BAĞLANIM DOĞRUSU



Örneklem Bağlanım İşlevi ve Rastsallık

- Anlaşılıyor ki rastsallık nedeniyle örneklem verilerini kullanarak anakütle bağlanım işlevini tam doğru biçimde tahmin etmek olanaksızdır.
- Elimizdeki iki değişik örneklem bağlanım doğrusundan hangisinin gerçek anakütle bağlanım doğrusunu daha iyi temsil ettiği kesin değildir.
- Genel olarak, n farklı örneklem için n sayıda farklı ÖBİ bulunabilir diyebiliriz.

Örneklem Bağlanım İşlevinin Bulunması

- Açıklamış olduğumuz tahmin sorunu yüzünden örneklem bağlanım işlevi aşağıdaki gibi gösterilir:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

- Burada:

$\hat{\beta}_1$ “ β_1 şapka” (β_1 hat) diye okunan β_1 ’in tahmincisini,
 $\hat{\beta}_2$ β_2 ’nin tahmincisini,
 \hat{u}_i u_i ’nin tahmincisini göstermektedir.

Örneklem Bağlanım İşlevinin Bulunması

- Anakütle bağlanım işlevini başta $Y_i = 17 + 0,6X_i + u_i$ olarak hesaplamış olduğumuzu anımsayalım.

- Bulduğumuz birinci örneklem bağlanım işlevi şudur:

$$Y_i = 24,5 + 0,509X_i + \hat{u}_i$$

- Bulduğumuz ikinci örneklem bağlanım işlevi ise şudur:

$$Y_i = 17,2 + 0,576X_i + \hat{u}_i$$

- Örneklem bağlanım işlevlerinin her ikisi de β_1 “değiştirge” (parameter) değerini yüksek tahmin ederken, β_2 değiştirge değerini düşük tahmin etmiştir.
- O zaman buradaki önemli soru, ABİ bilinemese bile $\hat{\beta}_1$ 'nin gerçek β_1 'e ve $\hat{\beta}_2$ 'nin da gerçek β_2 'ye olabildiğince yakın olduğu bir ÖBİ'nin nasıl oluşturulabileceği sorusudur.

Örneklem Bağlanım İşlevinin Bulunması

Gujarati'nin sözleriyle:

“Burada vurguladığımız, ABİ’yi olabildiğince doğru yansıtan ÖBİ’nin nasıl kurulacağını söyleyen süreçler geliştirebileceğimizdir. ABİ’yi asla gerçekten belirleyemesek bile, bunun yapılabileceğini düşünmek heyecan vericidir.”

- Bu heyecanlı tartışmayı burada şimdilik sonlandırıyoruz.

Önümüzdeki Dersin Konusu ve Ödev

Ödev

Kitaptan **Bölüm 1** “The Nature of Regression Analysis” ve **Bölüm 2** “Two-Variable Regression Analysis: Some Basic Ideas” okunacak.

Önümüzdeki Ders

İki Değişkenli Bağlanım Modeli: Tahmin Sorunu