

EKONOMETRİK MODELLER

PROGRAM: Gazi Üniversitesi FBE İstatistik YL

ÖĞRETİM ÜYESİ: PROF. DR. M. AKİF BAKIR

ÖDEV NO:5

İSİM SOYİSİM: GAMZE ZORLU

ÖĞRENCİ NO: 22830301032

DERS: EKONOMETRİK MODELLER
PROGRAM: Gazi Üniversitesi FBE İstatistik YL
ÖĞRETİM ÜYESİ: PROF. DR. M. AKİF BAKIR

ÖDEV 5:

Çoklu regresyon tahmini, Çoklu bağlantı irdelemesi

AMAÇ:

1. Bu ödevin amacı klasik çoklu doğrusal regresyon parametrelerinin EKK ile tahmin edilmesini ve çıktının yorumlanmasını öğrenmek.
2. Çoklu bağlantı problemini irdelemek

SON TARİH:

NOT: Çözümleri STATA'da yapınız.

LONGLEY VERİSİ

Aşağıda Longley tarafından toplanmış olan veriler yer almaktadır. Veriler orijinal olarak çeşitli bilgisayar programlarında EKK tahminlerinin hesaplama doğruluğunu değerlendirmek için toplanmış olmasına rağmen Longley verisi çoklu bağlantı probleminin de dâhil olduğu bazı ekonometrik problemleri göstermek üzere popüler bir veri haline gelmiştir. Veriler 1947-1962 arası zaman dizisi verisidir ve değişkenlerin tanımları şöyledir:

Y = istihdam edilen kişi sayısı, (x000);

X_1 = zımnî GSMH fiyat deflatörü,

X_2 = GSMH, (milyon \$);

X_3 = işsiz sayısı (x000),

X_4 = asker sayısı,

X_5 = 14 yaş üzeri kurumsal olmayan nüfus;

X_6 = yıl (1947=1, 1948=2, ve 1962=16).

TABLE 10.7 LONGLEY DATA

Observation	y	X_1	X_2	X_3	X_4	X_5	Time
1947	60,323	830	234,289	2356	1590	107,608	1
1948	61,122	885	259,426	2325	1456	108,632	2
1949	60,171	882	258,054	3682	1616	109,773	3
1950	61,187	895	284,599	3351	1650	110,929	4
1951	63,221	962	328,975	2099	3099	112,075	5
1952	63,639	981	346,999	1932	3594	113,270	6
1953	64,989	990	365,385	1870	3547	115,094	7
1954	63,761	1000	363,112	3578	3350	116,219	8
1955	66,019	1012	397,469	2904	3048	117,388	9
1956	67,857	1046	419,180	2822	2857	118,734	10
1957	68,169	1084	442,769	2936	2798	120,445	11
1958	66,513	1108	444,546	4681	2637	121,950	12
1959	68,655	1126	482,704	3813	2552	123,366	13
1960	69,564	1142	502,601	3931	2514	125,368	14
1961	69,331	1157	518,173	4806	2572	127,852	15
1962	70,551	1169	554,894	4007	2827	130,081	16

Source: See footnote 44.

Amacımız altı tane X değişkeninin hepsini kullanarak doğrusal modeli tahmin etmektir.

1. Doğrusal Regresyon modelinin parametrelerini tahmin ediniz. (Bilgisayar çıktısını düzenleyerek, okunabilir biçimde veriniz.)
2. Parametrelerin teker teker anlamlılık testlerini yapınız.
3. Modelin bütünü anlamlı mıdır?
4. Aynı modeli 1947-1961 dönemi verilerini kullanarak tahmin ediniz ve tüm 1947-1962 verisini kullanarak elde ettiğiniz tahmin modeliyle karşılaştırarak bu irdelemeden ulaştığınız sonucu açıklayınız.
5. Çoklu ortak doğrusallığın varlığını tüm öğretilen ölçütleri ayrı ayrı kullanarak araştırınız ve yorumlayınız.
6. Bu problem için ÇB problemi var ise bu problemi ortadan kaldıracak uygun yöntem/yöntemleri kullanarak analizi gerçekleştirip, sonuçları yorumlayınız. Yöntemi kullanma gerekçesini açıklayınız. (1)2de elde ettiğiniz modelle karşılaştırınız.

Soru-1)

. reg y x1 x2 x3 x4 x5 x6

Source	SS	df	MS	Number of obs	=	16
Model	184172402	6	30695400.3	F(6, 9)	=	330.29
Residual	836424.056	9	92936.0062	Prob > F	=	0.0000
				R-squared	=	0.9955
				Adj R-squared	=	0.9925
Total	185008826	15	12333921.7	Root MSE	=	304.85

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	15.06187	84.91493	0.18	0.863	-177.029 207.1528
x2	-.0358192	.033491	-1.07	0.313	-.1115811 .0399427
x3	-2.02023	.4883997	-4.14	0.003	-3.125067 -.915393
x4	-1.033227	.2142742	-4.82	0.001	-1.517949 -.548505
x5	-.0511041	.2260732	-0.23	0.826	-.5625172 .460309
x6	1829.151	455.4785	4.02	0.003	798.7875 2859.515
_cons	-3482259	890420.4	-3.91	0.004	-5496529 -1467988

model Denklemi :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6}$$

$$Y_i = -3482259 + 15.06187 X_{i1} - 0.0358192 X_{i2} - 2.02023 X_{i3} - 1.033227 X_{i4} - 0.0511041 X_{i5} + 1829.151 X_{i6}$$

Soru-2) $\beta_0 \Rightarrow p\text{-value} = 0,004$ + istatistik $t_i = -3.91$

$H_0: \beta_0 = 0$ (β_0 katsayısı model için anlamlı değildir)

$H_1: \beta_0 \neq 0$ (β_0 katsayısı model için anlamlıdır)

$p\text{-value} = 0,004 < \alpha = 0,05$ olduğundan H_0 hipotezi reddedilir.

Yani β_0 katsayısı %95 güven düzeyinde anlamlıdır.

$\beta_1 \Rightarrow p\text{-value} = 0,863$ + istatistik $t_i = 0,18$

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$p\text{-value} = 0,863 > \alpha$ olduğundan H_0 reddedilemez. Yani %95

güven düzeyinde β_1 katsayısı anlamlı değildir.

$\beta_2 \Rightarrow p\text{-value} = 0,313$ + istatistik $t_i = -1,07$

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

$p\text{-value} = 0,313 > \alpha$ olduğundan H_0 reddedilemez.

Yani %95 güven düzeyinde β_2 katsayısı model için anlamlı değildir.

$$\beta_3 \Rightarrow p\text{-value} = 0.003 \quad t\text{-istatistiği} = -4.14$$

$H_0: \beta_3 = 0$ $p\text{-value} = 0.003 < \alpha$ olduğundan H_0 reddedilir.

$H_1: \beta_3 \neq 0$ Yani %95 güven düzeyinde β_3 katsayısı model için anlamlıdır.

$$\beta_4 \Rightarrow p\text{-value} = 0.001 \quad t\text{-istatistiği} = -4.82$$

$H_0: \beta_4 = 0$ $p\text{-value} = 0.001 < \alpha$ olduğundan H_0 hipotezi reddedilir.

$H_1: \beta_4 \neq 0$ Yani β_4 katsayısı %95 güven düzeyinde model için anlamlıdır.

$$\beta_5 \Rightarrow p\text{-value} = 0.826 \quad t\text{-istatistiği} = -0.23$$

$H_0: \beta_5 = 0$ $p\text{-value} = 0.826 > \alpha$ olduğundan H_0 hipotezi reddedilmez.

$H_1: \beta_5 \neq 0$ Yani β_5 katsayısı %95 güven düzeyinde model için anlamlı değildir.

$$\beta_6 \Rightarrow p\text{-value} = 0.003 \quad t\text{-istatistiği} = 4.02$$

$H_0: \beta_6 = 0$ $p\text{-value} = 0.003 < \alpha$ olduğundan H_0 reddedilir. Yani

$H_1: \beta_6 \neq 0$ %95 güven düzeyinde β_6 katsayısı model için anlamlıdır.

Soru-3)

$$Y_i = -3482259 + 15.06187 X_1 - 0.0358192 X_2 - 2.02023 X_3 + 1.033227 X_4 - 0.0511041 X_5 + 1829.151 X_6$$

$$\text{std error: } (890420.4) \quad (84.91493) \quad (0.033491) \quad (0.4883997) \quad (0.2162342) \quad (0.2260732) \quad (455.4785)$$

$$t: \quad (-3.91) \quad (0.18) \quad (-1.07) \quad (-4.14) \quad (-4.82) \quad (-0.23) \quad (4.02)$$

$$R^2 = 0.9955 \quad sd = 9$$

$$p\text{-value} = 0.0000$$

H_0 : Model anlamlı değildir.

H_1 : Model anlamlıdır.

$p\text{-value} = 0.0000 < \alpha$ olduğundan H_0 hipotezi reddedilir. Yani %95 güven düzeyinde model anlamlıdır.

Bağımsız değişkenler bağımlı değişkeni varlığını %99.2 oranında açıklar.

4)

```
. drop if x6==1962
(1 observation deleted)
```

```
. reg y x1 x2 x3 x4 x5 x6
```

Source	SS	df	MS	Number of obs	=	15
Model	155088615	6	25848102.4	F(6, 8)	=	295.77
Residual	699138.24	8	87392.28	Prob > F	=	0.0000
				R-squared	=	0.9955
				Adj R-squared	=	0.9921*
Total	155787753	14	11127696.6	Root MSE	=	295.62

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	-20.51082	87.0974	-0.24	0.820	-221.3578 180.3361
x2	-.0273342	.0331748	-0.82	0.434	-.1038355 .0491671
x3	-1.952293	.4767006	-4.10	0.003*	-3.051567 -.8530199
x4	-.9582393	.2162271	-4.43	0.002*	-1.45686 -.4596187
x5	.0513397	.233968	0.22	0.832	-.4881915 .5908709
x6	1585.156	482.6832	3.28	0.011*	472.086 2698.225
_cons	-3017441	939728.1	-3.21	0.012*	-5184458 -850424.5

model denklemi:

$$Y_i = -3017441 - 20.51082x_1 - 0.0273342x_2 - 1.952293x_3 - 0.9582393x_4 + 0.0513397x_5 + 1585.156x_6$$

katsayıların anlamlılığına baktığımızda: $H_0: \beta_i = 0$
 $H_1: \beta_i \neq 0$

p-value $< \alpha$ ise H_0 red.

$\beta_0, \beta_3, \beta_4, \beta_6$ katsayıları model için anlamlıdır (% 95 güven düzeyi)

$\beta_1, \beta_2, \beta_5$ katsayıları model için anlamlı değildir

Modelin bütünü için; p-value = 0.0000 $R^2 = 0.9995$
Düzeltilmiş $R^2 = 0.9921$

Model anlamlıdır.

. ftest y x1 x2 x3 x4 x5 x6

*** Farrar-Glauber Multicollinearity Tests**

Ho: No Multicollinearity - Ha: Multicollinearity

* (1) Farrar-Glauber Multicollinearity Chi2-Test:
 Chi2 Test = 202.8242 P-Value > Chi2(15) 0.0000

* (2) Farrar-Glauber Multicollinearity F-Test:

Variable	F_Test	DF1	DF2	P_Value
x1	232.540	9.000	5.000	0.000
x2	2681.600	9.000	5.000	0.000
x3	56.201	9.000	5.000	0.000
x4	5.146	9.000	5.000	0.043
x5	623.878	9.000	5.000	0.000
x6	1341.810	9.000	5.000	0.000

* (3) Farrar-Glauber Multicollinearity t-Test:

Variable	x1	x2	x3	x4	x5	x6
x1
x2	26.534
x3	2.202	2.110
x4	1.593	1.549	-0.623	.	.	.
x5	16.217	20.738	2.743	1.199	.	.
x6	22.016	29.271	2.543	1.397	32.405	.

Ho: Çoklu bağlantı yoktur.
 H1: Çoklu bağlantı vardır.

p-value = 0.0000

$\chi^2 = 202.8242$

$p < \alpha$ olduğundan

Ho reddedilir.

Çoklu bağlantı vardır

. collin x1 x2 x3 x4 x5 x6
 (obs=15)

Collinearity Diagnostics

Variable	VIF	SQRT VIF	Tolerance	R- Squared
x1	130.19	11.41	0.0077	0.9923
x2	1490.72	38.61	0.0007	0.9993
x3	32.22	5.68	0.0310	0.9690
x4	3.86	1.96	0.2591	0.7409
x5	347.60	18.64	0.0029	0.9971
x6	746.46	27.32	0.0013	0.9987

Mean VIF 458.51

	Eigenval	Cond Index
1	6.8575	1.0000
2	0.0874	8.8559
3	0.0444	12.4284
4	0.0106	25.4752
5	0.0001	277.6460
6	0.0000	1111.0658
7	0.0000	45676.0390

Condition Number 45676.0390

Eigenvalues & Cond Index computed from scaled raw sscp (w/ intercept)

Det(correlation matrix) 0.0000

VIF = 458.51 > 10

Soru -5

```
. correlate x1 x2 x3 x4 x5 x6
(obs=16)
```

	x1	x2	x3	x4	x5	x6
x1	1.0000					
x2	0.9916	1.0000				
x3	0.6206	0.6043	1.0000			
x4	0.4647	0.4464	-0.1774	1.0000		
x5	0.9792	0.9911	0.6866	0.3644	1.0000	
x6	0.9911	0.9953	0.6683	0.4172	0.9940	1.0000

Bağımsız değişkenler arasında ikili ilişkiyi incelediğimizde genel olarak aralarında yüksek korelasyon olduğunu görüyoruz.

```
. vif
```

Variable	VIF	1/VIF
x2	1788.51	0.000559
x6	758.98	0.001318
x5	399.15	0.002505
x1	135.53	0.007378
x3	33.62	0.029745
x4	3.59	0.278635
Mean VIF	519.90	

VIF değeri bir bağımsız değişken diğer bağımsız değişkenlerle olan ilişkisinin derecesini belirlemek için hesaplanır. VIF değerinin 10'dan büyük olması çoklu bağlantı probleminin işaretlerinden biridir.

mean VIF = 519.90 > 10 olduğundan çoklu doğrusal bağlantı probleminin olduğuna işaret eder.

. fgtest y x1 x2 x3 x4 x5 x6

=====

*** Farrar-Glauber Multicollinearity Tests**

=====

Ho: No Multicollinearity - Ha: Multicollinearity

*** (1) Farrar-Glauber Multicollinearity Chi2-Test:**

Chi2 Test = 218.5559 P-Value > Chi2(15) 0.0000

*** (2) Farrar-Glauber Multicollinearity F-Test:**

Variable	F_Test	DF1	DF2	P_Value
x1	269.066	10.000	5.000	0.000
x2	3574.850	10.000	5.000	0.000
x3	65.238	10.000	5.000	0.000
x4	5.178	10.000	5.000	0.042
x5	796.307	10.000	5.000	0.000
x6	1515.957	10.000	5.000	0.000

*** (3) Farrar-Glauber Multicollinearity t-Test:**

Variable	x1	x2	x3	x4	x5	x6
x1	.					
x2	24.228	.				
x3	2.503	2.398	.			
x4	1.660	1.578	-0.570	.		
x5	15.248	23.530	2.986	1.237	.	
x6	23.610	32.409	2.841	1.452	28.624	.

Farrar - Glauber istatistikine göre, ortogonallikten sapma arttıkça çoklu doğrusal bağlantı durumu kuvvetlenir.

Ho: A veri matrisi ortogonaldır. (Çoklu bağlantı yoktur)

Hi: A veri matrisi ortogonal değildir. (Çoklu bağlantı vardır)

$$\chi^2 = 218.5559 \quad p\text{-value} = 0.0000$$

$p\text{-value} = 0.0000 < \alpha$ olduğundan Ho reddedilir.

Yani çoklu doğrusal bağlantı vardır

Soru - 6

Doğrusal Bağlantı Problemini Çözme Yolları

1) Gösteli sayısı artırma

2) Değişken Seçimi (Değişken silme)

Çoklu bağlantı problemini çözmek için bağımlı değişken üzerinde daha etkilili olan bağımsız değişken ya da değişkenler model dışında bırakılabilir.

3) Değişken Dönüştürme

Değişkenlerin farkını almak, logaritmasını almak ya da korelasyonu iki değişkeni birleştirilerek tek bir değişken olarak ifade etmek çoklu bağlantı problemini çözebilir ya da azaltabilir.

$$\text{Zimni GSMH fiyat deflatörü } (X_1) = \frac{\text{GSMH}(X_2)}{\text{reel GSMH}} \times 100$$

$$\text{reel GSMH} = \frac{X_2}{X_1} \times 100$$

4) Model seçimi yalnız olduğu iken olabilir. Model seçimi değiştirilebilir.

```
. gen reel_gsmh= ( x2 / x1)*100
```

```
. reg y reel_gsmh x3 x4 x5 x6
```

Source	SS	df	MS	Number of obs	=	16
Model	184031329	5	36806265.8	F(5, 10)	=	376.54
Residual	977497.183	10	97749.7183	Prob > F	=	0.0000
				R-squared	=	0.9947
				Adj R-squared	=	0.9921
Total	185008826	15	12333921.7	Root MSE	=	312.65

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
reel_gsmh	-.0067896	.02341	-0.29	0.778	-.0589503	.0453711
x3	-1.580112	.3169032	-4.99	0.001	-2.286216	-.8740075
x4	-.9288877	.1946236	-4.77	0.001	-1.362536	-.4952393
x5	-.2173108	.1291503	-1.68	0.123	-.5050757	.0704541
x6	1379.172	325.3048	4.24	0.002	654.3479	2103.996
_cons	-2594741	623291.5	-4.16	0.002	-3983521	-1205961

```
. vif
```

Variable	VIF	1/VIF
x6	368.08	0.002717
reel_gsmh	292.70	0.003416
x5	123.85	0.008074
x3	13.46	0.074310
x4	2.82	0.355234
Mean VIF	160.18	

X_1 ve X_2 yi $\frac{X_2}{X_1} \times 100$ olarak tek bir değişken olarak ifade edersek ve sonrasında model kurarsak çoklu bağlantı probleminin $VIF \approx 160.18$ olarak çıktığını fakat gözden geçirilmediğini görürüz.

$X_5 = 14$ yaş üzeri kurumsal olmayan nüfus, $X_6 = 41$ yaş üstü nüfus
arttığından X_6 yi model dışında bırakıp deneysek;

. reg y reel_gsmh x3 x4 x5

Source	SS	df	MS	Number of obs	=	16
Model	182274330	4	45568582.4	F(4, 11)	=	183.31
Residual	2734496.29	11	248590.572	Prob > F	=	0.0000
				R-squared	=	0.9852
				Adj R-squared	=	0.9798
Total	185008826	15	12333921.7	Root MSE	=	498.59

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
reel_gsmh	.0739538	.0217092	3.41	0.006	.0261723 .1217354
x3	-.5359239	.318017	-1.69	0.120	-1.235875 .1640267
x4	-.7272152	.3009571	-2.42	0.034	-1.389617 -.064813
x5	-.0540364	.1965889	-0.27	0.789	-.4867256 .3786527
_cons	47497.85	14656.23	3.24	0.008	15239.7 79756

. vif

Variable	VIF	1/VIF
x5	112.84	0.008862
reel_gsmh	98.98	0.010103
x3	5.33	0.187658
x4	2.65	0.377803
Mean VIF	54.95	

Farrar-Glauber Chi2 Test = 75.5080
p-value = 0.0000

$X_6 = 4.1$ değışkeni modele dahil edilmediğinde VIF değeri düşmüř fakat çoklu bağımlılık problemi piderilmekmiştir.

Log-Log, Lin-Log, Log-Lin modelleri uygulanmış fakat çoklu bağımlılık problemi bu modellerde de mevcuttur.