SPSS İLE İSTATİSTİKSEL VERİ ANALİZİ

Statistical Packages for the Social Sciences

PROF.DR.YÜKSEL TERZİ

ONDOKUZ MAYIS ÜNİVERSİTESİ
FEN-EDEBİYAT FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ
SAMSUN

2019

ÇOKLU DOĞRUSAL REGRESYON

Çoklu regresyon çözümlemesinde bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi matematiksel modelle açıklayarak, bağıntılar bulmak ve bağımsız değişkenler yardımı ile bağımlı değişkenin kestirimi yapılır.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

Modelde amaç;

- Bağımsız değişkenler yardımı ile bağımlı değişkeni kestirmek,
- Bağımsız değişkenlerden hangisi-hangilerinin bağımlı değişkeni daha çok etkilediğini bulmak ve aralarındaki karmaşık yapıyı tanımlamaktır.

$$\mathrm{E}(Y|X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

$$H_1 : En\,az\,bir\;\beta_j \neq 0$$

Çoklu Regresyon Analizi Varsayımları

a) Bağımlı değişken bir tesadüfi bir değişkendir ve normal dağılım göstermektedir.

b) Tahmin hataları tesadüfidir, birbiri ile ilişki göstermezler ($Cov(\varepsilon_i, \varepsilon_j) = 0$) ve ortalaması sıfır ($E(\varepsilon) = 0$) ve varyansı $V(\varepsilon) = \sigma^2 I$ şeklindedir. Bu matris varyans kovaryans matrisidir. Bu matrisin köşegen elemanları σ^2 olup diğer elemanları sıfırdır.

c) Bağımsız değişkenler rassal değişkenler değildir. Yani bağımsız değişkenlerin aldıkları değerler sabit veya önceden belirlenmiş yada isteğe bağlı olarak seçilmiş değerler olduğu varsayılır.

d) Hatalar birbirinden bağımsızdır (**otokorelasyon yoktur.)**

e) Her bağımsız değişkenin değerlerine ait olan bağımlı değişken değerlerinin alt setleri varyansları birbirine eşittir (**esit varyanslılık=homoscedasticity).**

f) Bağımsız değişkenler arasında basit doğrusal ilişkilerin olmaması gerekir. Bağımsız değişkenler arasındaki basit doğrusal korelasyon katsayılarının 0 veya 0'a çok yakın olması şartı seklinde de açıklanabilen bu varsayıma, istatistikte **"Çoklu Doğrusal Bağlantı" (Multicollinearity)** olmama durumu adı verilmektedir.

g) Bağımsız değişkenler ile hata terimi arasında ilişki yoktur.

$Cov(\varepsilon_i, X_{ji}) = 0$, j=1,2,.....,p ve i=1,2,...,n

h) Hipotez testlerini yapabilmek ve güven aralıklarını oluşturabilmek için hataların normal dağılıma sahip olduğu varsayılır.

EKK TAHMİN EDİCİLERİN VARSAYIMLARI (Gauss-Markov Teoremi)

1. Parametrelerde doğrusallık (Linear in parameters)
2. Rassal örnekleme (random sampling)
3. Sıfır koşullu ortalama (Zero conditional mean)

Açıklayıcı değişkenlerle hatalar ilişkisizidir. $E(e|X_1,X_2,\ldots,X_k)=0$

Eğer bu varsayım sağlanmazsa;

i) Regresyonun fonksiyonel biçimi yanlış seçilmiştir.
ii) Önemli bir değişken model dışında bırakılmıştır.
iii) Değişkenin ölçümünde hatalar vardır.

4. Çoklu bağlantı olmaması (No perfect collinearity)

Açıklayıcı değişkenler arasında yüksek korelasyon olmaması gerekir. X'lerden birisi diğer X'lerin doğrusal kombinasyonu durumunda ise çoklu bağlantı durumu ortaya çıkar.

$Y=b_0+b_1X_1+b_2X_2+b_3(X_1+X_2)+e$, $\log Y=b_0+b1\log X+b_2\log X^2+e$

Ancak $Y=b_0+b_1X+b_2X^2+e$ modeli çoklu bağlantıya sebep olmaz.

5. Sabit varyans (hemoscedasticity)

Artık terimlerinin varyansı sabit olduğu varsayılır. Böylece varyans formülleri basit hale gelir ve tahmin ediciler etkinlik özelliği kazanır.

$$V(e|X_1,\ldots,X_k)=\sigma^2$$

Yukarıdaki 5 varsayım sağlanırsa Gauss-Markov teoremine göre EKK tahmin edicileri sapmasız ve minimum varyanslıdırlar.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

β_2: Sabit, hata ve X1 sabitken X_2'deki 1 birimlik değişmenin Y'de yapacağı değişmeyi gösterir.

$$\log Y = \beta_0 + \beta_1 \log X_1 + \beta_2 X_2 + \varepsilon$$

β_1: Sabit, hata ve X_2 sabitken X_1'deki %1'lik artışın Y'de yapacağı % artışı gösterir.

$$\mathrm{Y} = \mathrm{X}\beta + \varepsilon$$

$$\mathrm{X}'\mathrm{X}\hat{\beta} = \mathrm{X}'\mathrm{Y}$$

$$X'Y = \begin{bmatrix} \sum Y_i \\ \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \\ \vdots \\ \sum X_{pi}Y_i \end{bmatrix}$$

$$\hat{\beta} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{Y}$$

Çoklu Regresyonda ANOVA Tablosu

Değişim Kaynağı	sd	KT	KO	F
Regresyon	p	$\hat{\beta}'X'Y - n\overline{Y}^2$	RKO=RKT/p	RKO/AKO
Artık	n-p-1	$Y'Y - \hat{\beta}'X'Y$	AKO=AKT/(n-p-1)	
Genel	n-1	$Y'Y - n\overline{Y}^2$		

$$R^2 = \frac{RKT}{GKT} = \frac{\hat{\beta}'X'Y - n\overline{Y}^2}{Y'Y - \hat{\beta}'X'Y}$$

Model	df	Type I (Sequential) Sum of Squares	Type III (Partial) Sum of Squares
Regression	3		
X_1	1	$SS(X_1)$	$SS_1(X_1 \mid X_2,X_3)=SS(X_1,X_2,X_3)-SS(X_2,X_3)$
X_2	1	$SS(X_2 \mid X_1)=SS(X_1,X_2)-SS(X_1)$	$SS_2(X_2 \mid X_1,X_3)=SS(X_1,X_2,X_3)-SS(X_1,X_3)$
X_3	1	$SS(X_3 \mid X_1,X_2)=SS(X_1,X_2,X_3)-SS(X_1,X_2)$	$SS_3(X_3 \mid X_2,X_1)=SS(X_1,X_2,X_3)-SS(X_2,X_1)$
Residual	n-k-1		
Total	n-1		

Çoklu Korelâsyon Katsayısı

Regresyonda bağımlı değişenin gözlem değerleri ile tahmin değerleri arasındaki Pearson korelasyon katsayısına çoklu korelasyon katsayısı denir. Bağımsız değişkenler ile bağımlı değişken arasındaki ne derece ilişki olduğunu gösteren ölçüdür.

$$R_{Y\hat{Y}} = \frac{\sum\left(Y_i - \overline{Y}\right)\left(\hat{Y}_i - \overline{Y}\right)}{\sqrt{\sum\left(Y_i - \overline{Y}\right)^2 \sum\left(\hat{Y}_i - \overline{Y}\right)^2}} \qquad \left(R_{Y\hat{Y}}\right)^2 = R^2$$

Kısmi Korelâsyon Katsayısı

Y ve X_1 değişkenleri arasındaki korelasyon katsayısını hesaplarken X_2 yada herhangi bir üçüncü değişkenin etkisinin hem Y hem de X_1 üzerinden arındırılması istenebilir. Üçüncü bir değişkenle hem Y hem de X_1 arasında ilişki olabilir. Böyle durumlarda üçüncü değişkenin etkisini hem Y hem de X_1 üzerinden arındırıp yeni elde edilecek değişkenler arasındaki korelasyon katsayısı hesaplanabilir. Bu tür korelasyon katsayılarına kısmi korelasyon katsayısı adı verilir. Kısmi korelasyon $(X_3, X_4, \ldots, X_p)$ değişkenlerinin etkisi X_1 ve X_2'den arındırıldıktan sonra bu iki değişken arasındaki doğrusal ilişkinin derecesini ölçen korelasyon katsayısıdır.

Zero-order (r) **:** Her bir bağımsız değişkenin bağımlı değişken(Y) ile korelasyonu

Partial (kısmi) **:** Diğer tüm bağımsız değişkenlerin X_i ve Y üzerindeki etkisini arıttıktan sonra bağımsız değişken ile Y arasındaki korelasyon katsayısıdır.

$\mathrm{r}_{\mathrm{YX}_1.\mathrm{X}_2}$ **:** X_2 değişkeninin etkisi hem Y hem de X_1 üzerinden arındırıldıktan sonra yeni elde edilen değişkenler arasındaki korelasyon katsayısı (kısmi korelasyon katsayısı)

$$r_{YX_1.X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{(1 - r^2_{YX_2})(1 - r^2_{X_1X_2})}}$$

Kısım (Part) Korelâsyon Katsayısı (Part or Semi-Partial Correlation)

Çoklu doğrusal regresyonda modele yeni bağımsız değişkenler ilave ederken bağımsız değişkenlerin kendi aralarında sıkı ilişkili olmasına dikkat edilmelidir. Eğer r_{X1X2}=0,99 ise modeldeki ikinci bağımsız değişkenin belirleme katsayısı üzerine etkisi hemen hemen sıfırdır.

Kısım korelasyon katsayısı diğer tüm bağımsız değişkenlerin X_i değişkeni üzerindeki etkisini arıttıktan sonra Y ile arasındaki korelasyon katsayısıdır.

Kısım korelasyon katsayısı o açıklayıcı değişkenin belirleme katsayısına olan net katkısı hesaplanabilir.

$$r_{Y(X_1.X_2)} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{1 - r^2_{X_1X_2}}}$$

$r_{Y(X_1.X_2)}$: X_2 değişkeninin etkisi X_1 üzerinden arındırıldıktan sonra elde edilen yeni değişkenle Y arasındaki kısmi korelasyon katsayısı

ÇOKLU REGRESYONDA MODELE GİRECEK DEĞİŞKENLERİN SEÇİMİ

Ölçüm yapılan bağımsız değişkenle, bağımlı değişken arasındaki ilişki, değişken sayısı arttıkça daha iyi izah edilir duruma gelir. Ancak, değişken sayısının arttırılması ek ölçümleri gerektirdiğinden zahmetli ve masraflı bir iştir. Bu nedenle toplam varyansı en az sayıda değişkenle açıklamak esas amaçtır.

Modele eklenmesi ile, bağımlı değişkenin varyasyonunu açıklamada önemli artış sağlayan değişkenleri belirlemek veya seçmek için değişik yöntemler vardır. Değişken seçimi üç veya daha fazla bağımsız değişken olduğu durumlarda önem kazanmaktadır.Değişken seçiminde sıkça kullanılan yöntemler aşağıdaki gibidir.

1) Mümkün Olan Tüm Regresyonlar (All Possible Regression-Enter)
2) Değişken Ekleme İşlemi (Forward Selection)
3) Değişken Eleme İşlemi (Backward Selection)
4) Değişken Ekleme ve Eleme İşlemi (Stepwise Selection)

1. Mümkün Olan Tüm Regresyonlar (All Possible Regression)

Burada amaç, en iyi alt regresyon setini bulmaktır. Yani Y'yi (bağımlı değişkeni) ifade edebilecek en az sayıda bağımsız X değişkeni belirlenecektir. Mümkün olan tüm regresyonlar yönteminde, ilk olarak değişkenlerin tekli, sonra ikili kombinasyonları, sonra üçlü kombinasyonları, ve n'inci kombinasyonlarını içeren modeller (regresyon setleri) göz önüne alınır. Öncelikle bir ön eleme yapılarak, bunlar arasında R^2'si (belirleme katsayısı) en yüksek olan regresyon setleri seçilir. Daha sonra bu setlerdeki değişkenlerin her birinin sırası ile modelde olmaması durumunda (hata kareler toplamı kullanılarak), modele katkısının önemli olup olmadığı kontrol edilir. Modele katkısı önemli olmayan değişkenler modelden atılır ve en iyi regresyon seti bulunana kadar işleme devam edilir.

$$F_c = \frac{R_p^2 - R_{p-i}^2}{\frac{1 - R_p^2}{n-p-1}} \sim F_{1,n\text{-}p\text{-}1,\alpha}$$

R_p^2 = Eşitlikte p tane terim olduğunda elde edilen (tüm modele ait) determinasyon katsayısını,

R_{p-1}^2 = Eşitlikten terim atıldıktan sonra elde edilen (indirgenmiş modelin) determinasyon katsayısını,

$F_{1,n\text{-}p\text{-}1,\alpha}$ = Tüm regresyonun F cetvel değerini

n - p -1= Tüm regresyonun hata serbestlik derecesini,

ifade eder.

R_p^2 değeri, R_0^2 değerinden yüksek olan setler önemli setlerdir.

R_0^2 değeri,

$$R_0^2 = 1 - (1 - R_t^2).(1 + \frac{t.f_{t(n-p-1)}}{n-p-1})$$

şeklindedir. Burada,

R_t^2 =Tüm regresyonun R^2'sini,

$f_{t(n-p-1)}$ = F cetvel değerini,

t= Regresyondaki bağımsız değişken sayısını,

n - p -1= Tüm regresyonun hata serbestlik derecesini,

ifade eder.

Fc hesap değeri, F cetvel değerinden küçük ise değişkenin modelde kalmasının modele fazla bir katkı sağlamayacağına karar verilir ve model dışı bırakılır.

2. Değişken Ekleme İşlemi (Forward Selection)

Öncelikle en iyi tek değişkenli model ile başlanarak her seferinde en yüksek katkıyı sağlayacak değişken ilave edilerek uygulanan bir yöntemdir. Yani bir değişken ilavesi ile, regresyonun hata kareler toplamında meydana gelen değişmenin (küçülmenin) önemli olup olmadığına bakılır. Bu işlem eklenen değişkenin katsayısının önemsiz olması durumunda sona erdirilir. Test istatistiği,

$$F_i = Max_i\left[\frac{(HKT_p - HKT_{p-i}}{\hat{\sigma}^2_{p-i}}\right] > F_c$$

şeklindedir. Burada,

HKT_p = Değişken eklenmesi durumunda oluşan regresyon setinin hata kareler toplamını,

HKT_{p-i} = Değişken ilave edilmeden önceki regresyon setine ait hata kareler toplamını,

$\hat{\sigma}^2_{p-i}$ = Değişken ilave edilmeden önceki regresyon setine ait varyansı, ifade eder.

3. Değişken Eleme İşlemi (Backward Selection)

İlk aşamada model içine tüm değişkenler dahil edilir. Daha sonraki kademelerde her defasında bir tane olmak üzere en düşük kısmi F değerine sahip olan bağımsız değişken (X) atılmak sureti ile işleme devan edilir. Atılan değişkenin katkısı her seferinde test edilir. Atılan değişkenin katkısı istatistiki olarak önemli ise atma işlemi gerçekleştirilmez ve işlem orada durdurulur.

$$F_i = \mathrm{Min}_i\left[\frac{(HKT_{p-i} - HKT_p}{\hat{\sigma}^2_p}\right] > F_c$$

HKT_p = Değişken çıkarılmadan önceki regresyon setine ait hata kareler toplamını,

HKT_{p-i} = Değişken çıkarılması durumunda oluşan regresyon setinin hata kareler toplamını,

$\hat{\sigma}^2_{p-i}$ = Değişken çıkarılmadan önceki durumunda oluşan regresyon setinin varyansını,

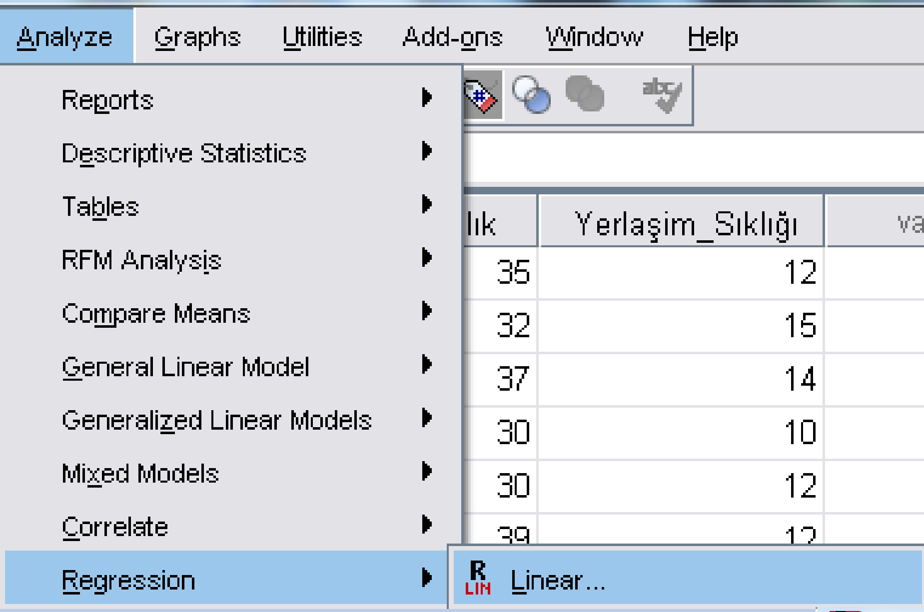
ifade eder.

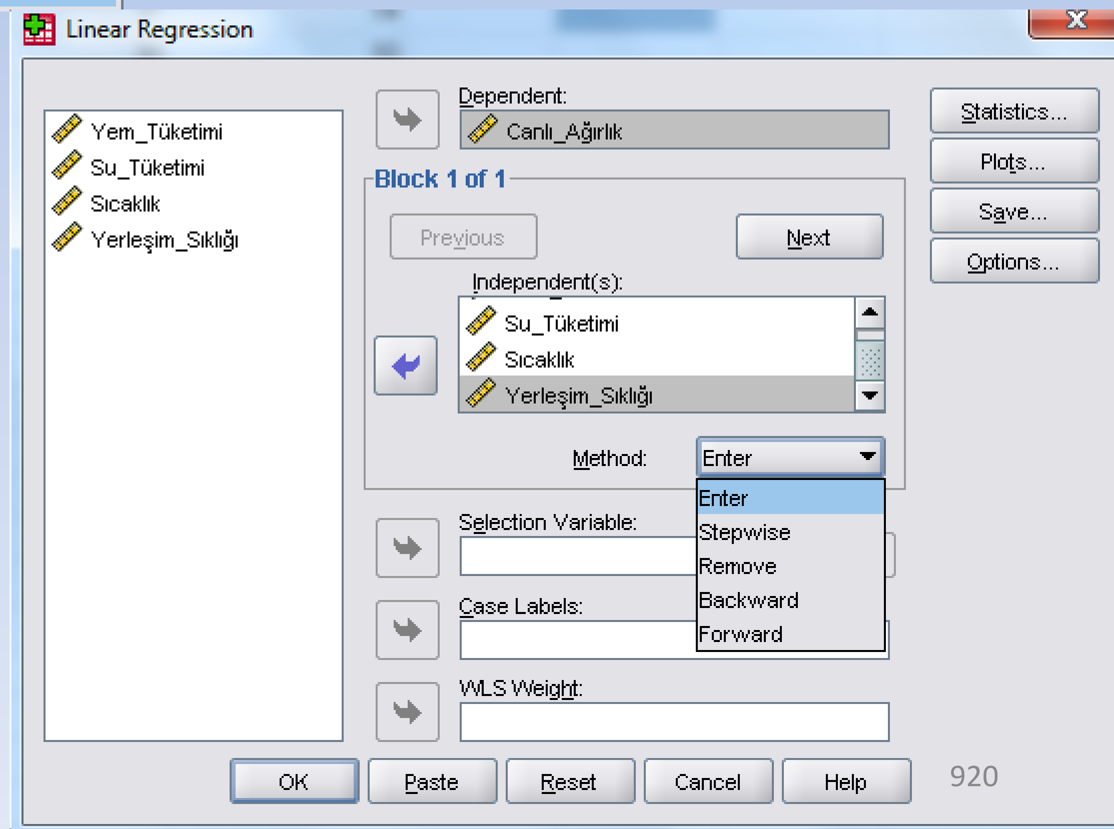
4. Değişken Ekleme Ve Eleme İşlemi (Stepwise Selection)

Esas olarak değişken ekleme yöntemine benzemektedir. Ancak bir değişken modele alındıktan sonra eleme yöntemine benzer şekilde modeldeki tüm değişkenlerin durumu yeniden incelenir. Bu yönüyle eleme yöntemine benzemektedir. Bu yöntemde işlem, herhangi bir değişken ilave edilemez veya atılamaz hale geldiğinde durdurulur. ***Yöntemin iyi tarafı daha önce modele girmiş bir değişken daha sonra modele girecek değişkenlerle ilişkilerine bağlı olarak atılabilir hale de gelmesidir.***

Örnek. Etlik piliçlerde yem tüketimi (gr/gün), su tüketimi (gr/gün), sıcaklık (^0C) ve yerleşim sıklığının (piliç/m^2) günlük **canlı ağırlık artışına (Y)** etkisi araştırılıyor. Elde edilen veriler aşağıdaki gibidir.

	Canlı_Ağırlık	Yem_Tüketimi	Su_Tüketimi	Sıcaklık	Yerlaşim_Sıklığı
1	34	110	225	35	12
2	39	112	251	32	15
3	41	136	235	37	14
4	30	109	190	30	10
5	30	108	162	30	12
6	36	98	198	39	12
7	31	91	154	39	9
8	44	133	260	28	16
9	43	127	210	28	17
10	39	114	154	27	15
11	35	107	188	25	11
12	45	139	245	27	14
13	45	141	229	25	15
14	43	129	212	28	17
15	40	116	156	27	15
16	37	109	190	25	13
17	45	141	248	27	15
18	46	143	231	25	16


Analyze Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
Tables
RFM Analysis
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Linear...
lık
Yerlaşim_Sıklığı
35 12
32 15
37 14
30 10
30 12


Linear Regression
Yem_Tüketimi
Su_Tüketimi
Sıcaklık
Yerleşim_Sıklığı
Dependent:
Canlı_Ağırlık
Block 1 of 1
Previous
Next
Independent(s):
Su_Tüketimi
Sıcaklık
Yerleşim_Sıklığı
Method:
Enter
Enter
Stepwise
Remove
Backward
Forward
Selection Variable:
Case Labels:
WLS Weight:
Statistics...
Plots...
Save...
Options...
OK
Paste
Reset
Cancel
Help

Enter Yönt.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2,214	7,563		,293	,774
	Yem_Tüketimi	,158	,064	,470	2,476	,028
	Su_Tüketimi	,016	,022	,104	,724	,482
	Sıcaklık	-,006	,139	-,005	-,041	,968
	Yerleşim_Sıklığı	1,072	,332	,463	3,232	,007

a. Dependent Variable: Canlı_Ağırlık

Forward Yönt.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3,384	4,761		,711	,488
	Yem_Tüketimi	,297	,039	,884	7,556	,000
2	(Constant)	2,368	3,701		,640	,532
	Yem_Tüketimi	,184	,045	,548	4,103	,001
	Yerleşim_Sıklığı	1,056	,309	,456	3,413	,004

a. Dependent Variable: Canlı_Ağırlık

Backward Yönt.

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2,214	7,563		,293	,774
	Yem_Tüketimi	,158	,064	,470	2,476	,028
	Su_Tüketimi	,016	,022	,104	,724	,482
	Sıcaklık	-,006	,139	-,005	-,041	,968
	Yerleşim_Sıklığı	1,072	,332	,463	3,232	,007
2	(Constant)	1,947	3,784		,514	,615
	Yem_Tüketimi	,159	,055	,473	2,866	,012
	Su_Tüketimi	,016	,019	,101	,800	,437
	Yerleşim_Sıklığı	1,074	,314	,464	3,421	,004
3	(Constant)	2,368	3,701		,640	,532
	Yem_Tüketimi	,184	,045	,548	4,103	,001
	Yerleşim_Sıklığı	1,056	,309	,456	3,413	,004

a. Dependent Variable: Canlı_Ağırlık

Stepwise Yönt.

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3,384	4,761		,711	,488
	Yem_Tüketimi	,297	,039	,884	7,556	,000
2	(Constant)	2,368	3,701		,640	,532
	Yem_Tüketimi	,184	,045	,548	4,103	,001
	Yerleşim_Sıklığı	1,056	,309	,456	3,413	,004

a. Dependent Variable: Canlı_Ağırlık

Remowe Yönt.

Coefficients[a]

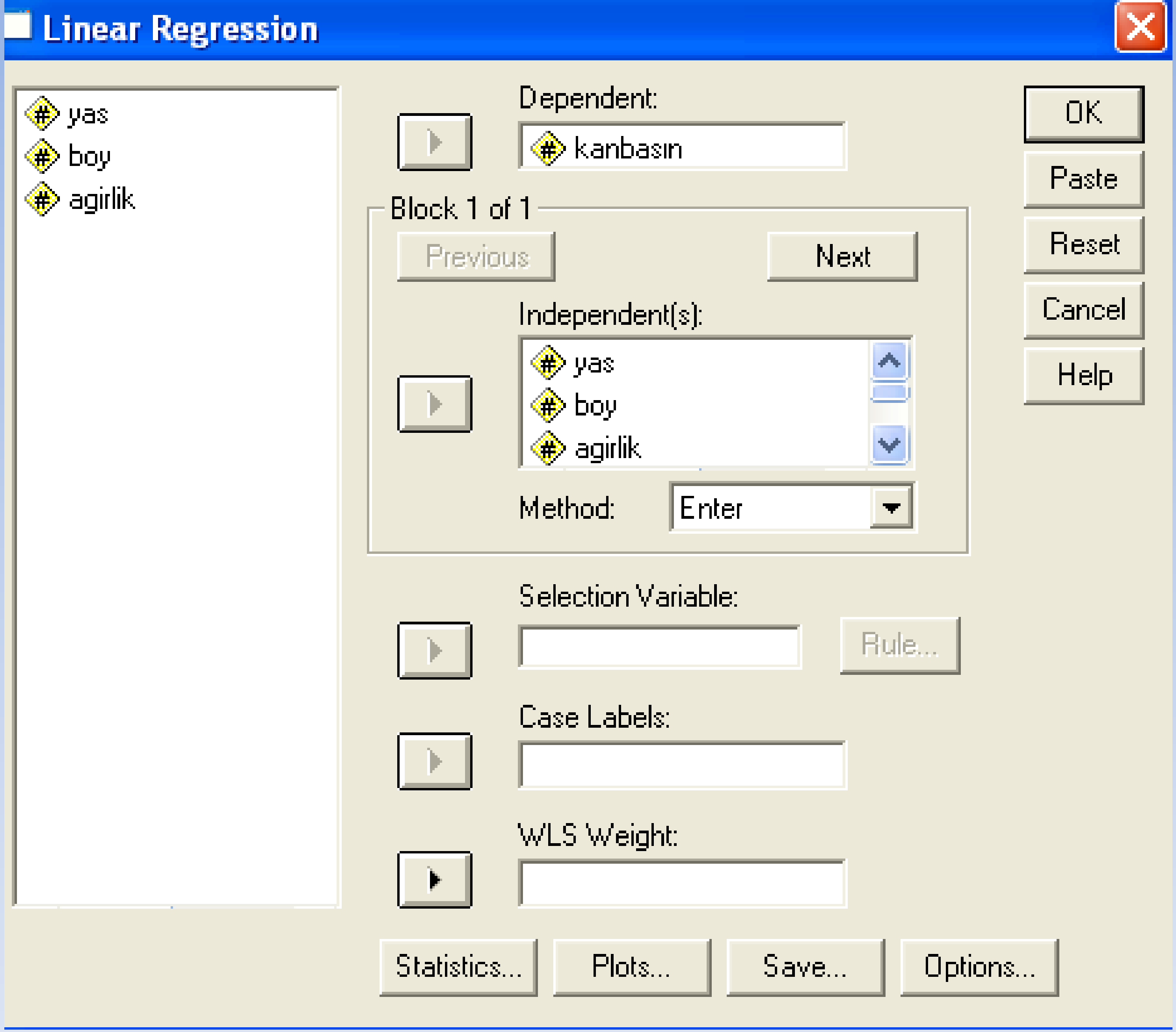
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2,214	7,563		,293	,774
	Yem_Tüketimi	,158	,064	,470	2,476	,028
	Su_Tüketimi	,016	,022	,104	,724	,482
	Sıcaklık	-,006	,139	-,005	-,041	,968
	Yerleşim_Sıklığı	1,072	,332	,463	3,232	,007
2	(Constant)	39,056	1,277		30,586	,000

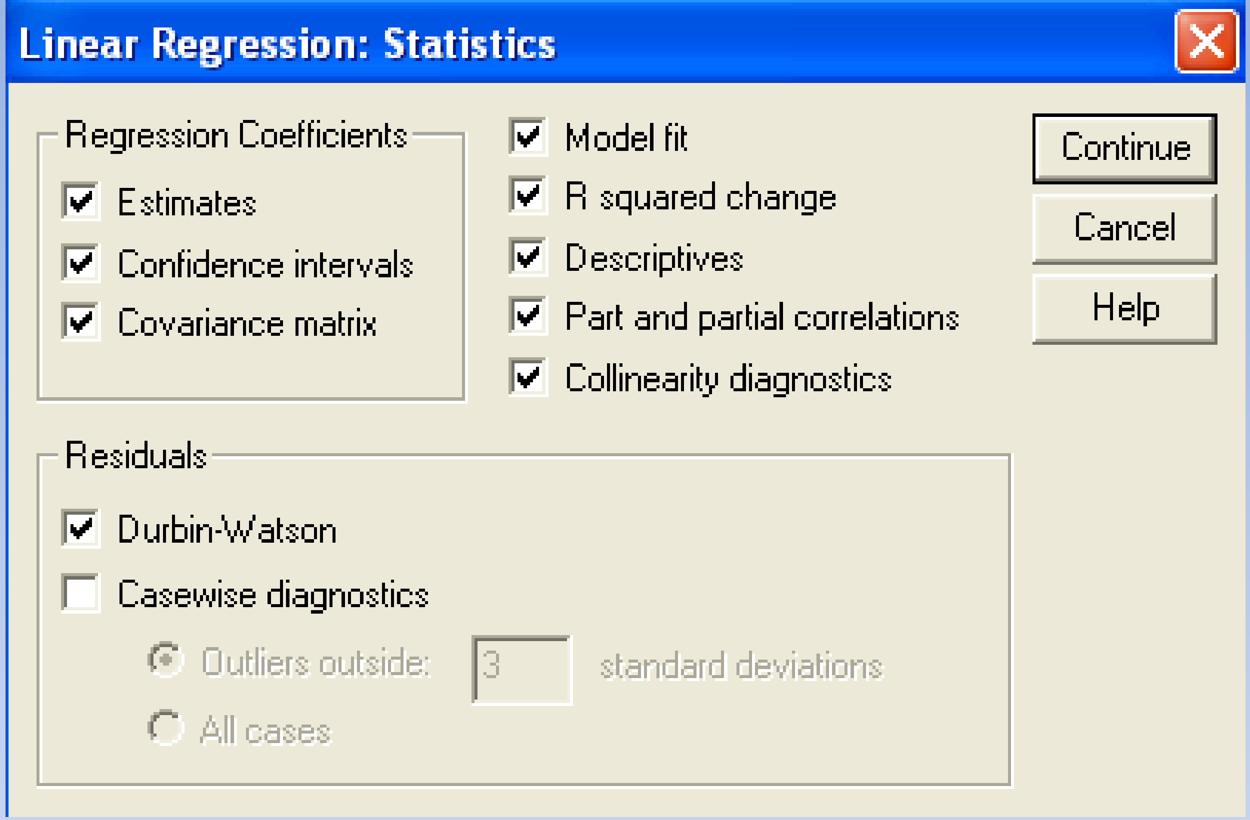
a. Dependent Variable: Canlı_Ağırlık

Örnek.

İnsanların kan basıncı üzerinde yaş,boy ve ağırlıklarının etkisi araştırılmış ve yandaki veriler elde edilmiştir. Kan basıncı üzerinde hangi açıklayıcı değişken(lerin) önemli olduğunu test edininiz.

	yas	boy	agirlik	kanbasın
1	51	166	67,00	115
2	64	165	61,00	122
3	46	174	83,00	130
4	39	168	78,90	126
5	58	162	67,00	110
6	54	178	90,00	141
7	31	171	77,70	124
8	67	173	89,30	150
9	48	165	70,00	110
10	39	177	82,50	130
11	51	166	63,00	120
12	73	178	93,10	149
13	56	169	72,00	125
14	47	159	64,00	114

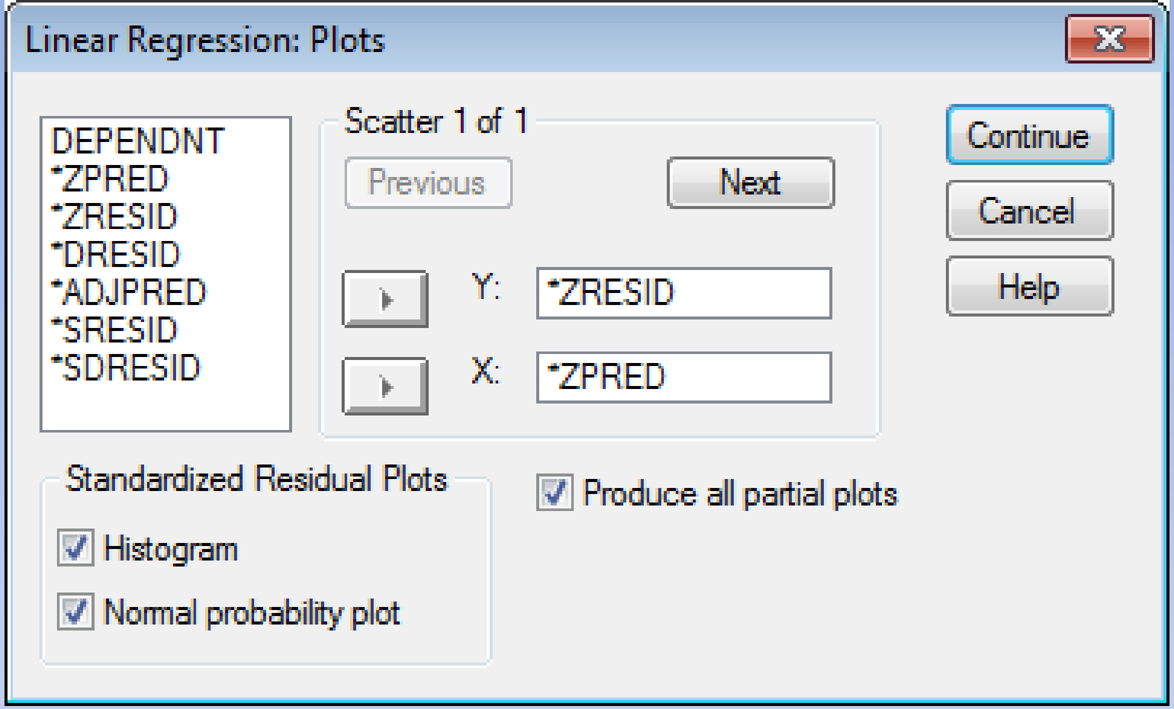

Linear Regression
yas
boy
agirlik
Dependent:
kanbasın
Block 1 of 1
Previous
Next
Independent(s):
yas
boy
agirlik
Method:
Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
OK
Paste
Reset
Cancel
Help
Statistics...
Plots...
Save...
Options...



Collinearity Diagnostics : **Bağımsız değişkenler arasında doğrusal bir ilişki olup olmadığını araştırır.**

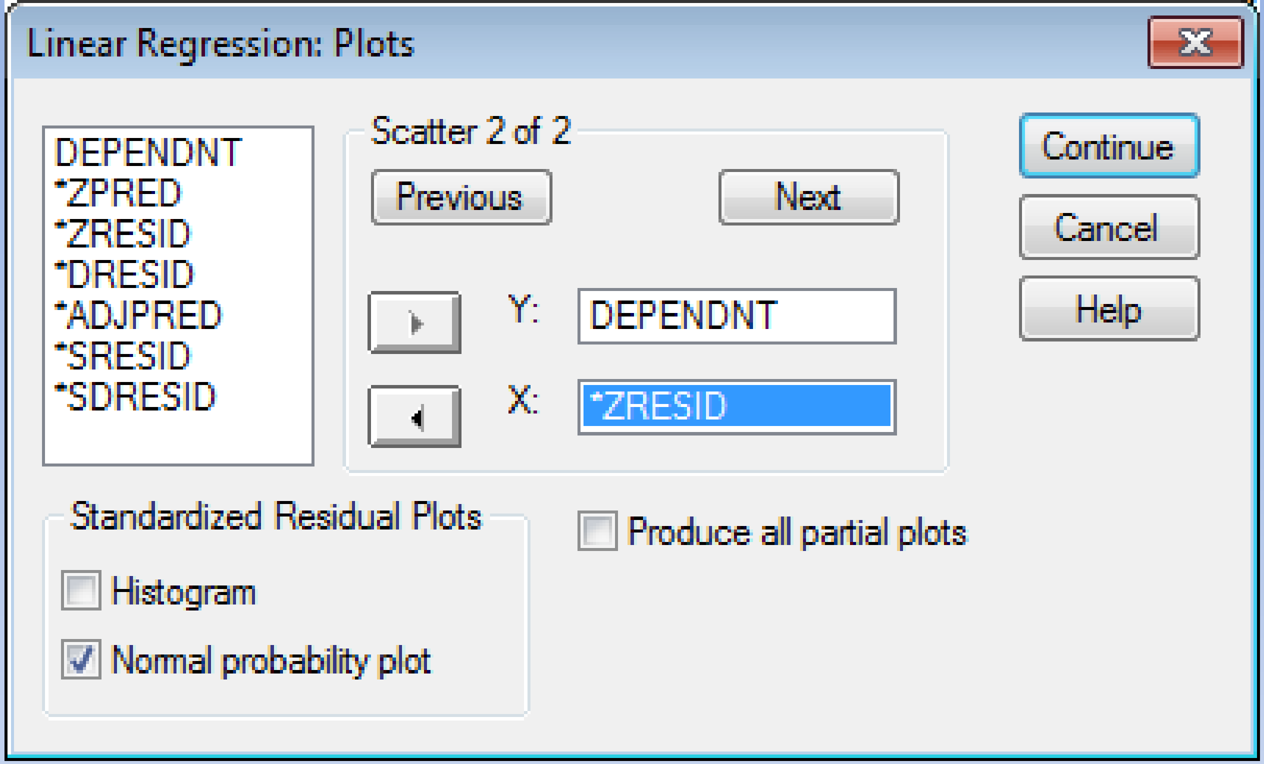
Model Fit : **Modele eklenen ve modelden çıkarılan değişkenler incelenir. R^2 ler ve varyans tablosu analiz edilir.**

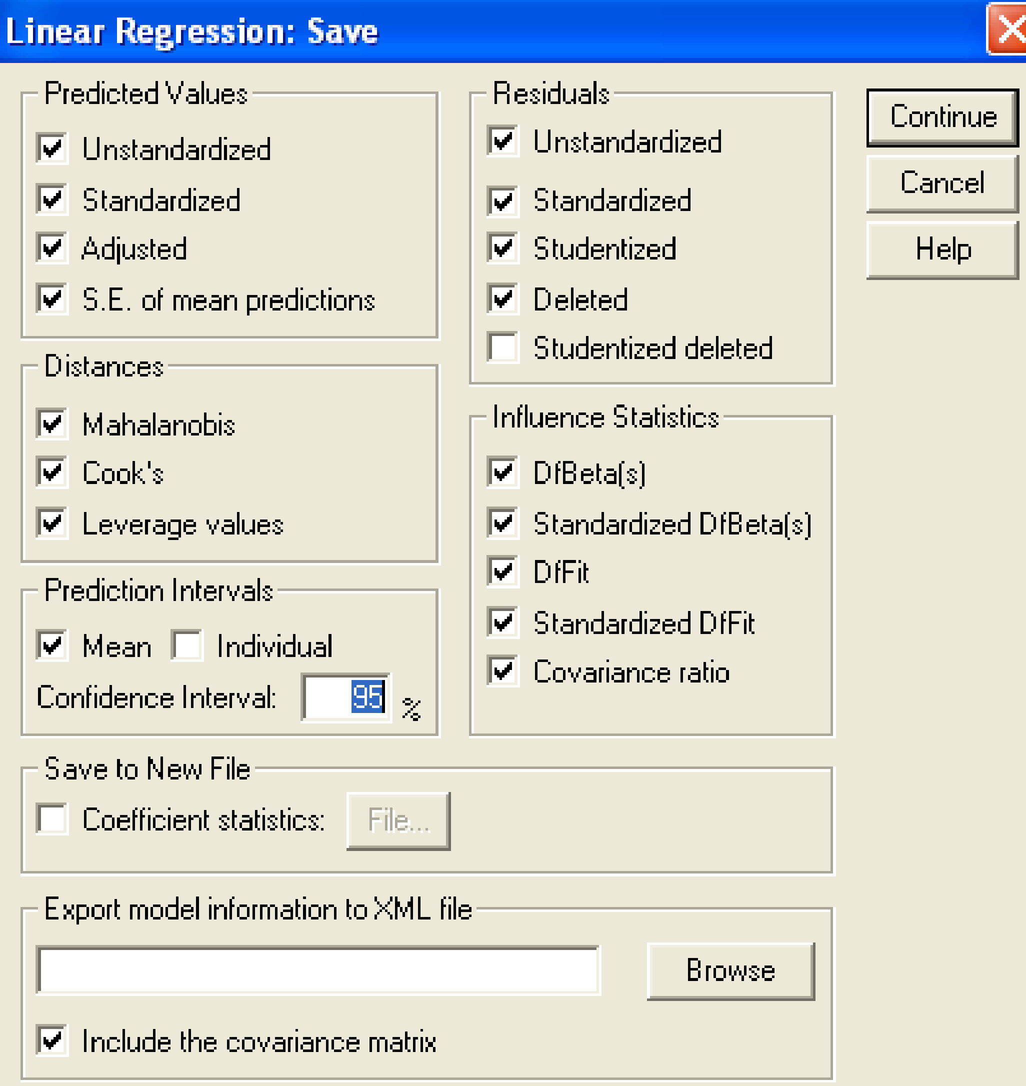
Durbin Watson : **Otokorelasyonu test eder. 0-4 arasında değerler alır. DW değeri 1.5-2.5 arasında ise otokorelasyon olmadığını gösterir. 2'ye yakın olması arzu edilir. 0'a yakın ise pozitif korelasyonu yani b katsayılarının s.hatalarının çok küçük olduğunu, 4'e yakın değerler ise negatif korelasyonu yani b katsayılarının s.hatalarının çok büyük olduğu anlamına gelir.**



ZPRED : **Standartlaştırılmış tahmini değerler**
ZRESID : **Standartlaştırılmış artıklar (residual)**
DRESID : **Silinen artıklar**
ADJPRED : **Düzeltilmiş tahmini değerler**
SRESID : **Student artıklar**
SDRESID : **Student silinen artıklar**

Histogram ve normal probability plot kısmından çoklu normal dağılım ve doğrusallık varsayımı kontrol edilebilir.


Linear Regression: Plots
DEPENDNT
*ZPRED
*ZRESID
*DRESID
*ADJPRED
*SRESID
*SDRESID
Scatter 2 of 2
Previous
Next
Y:
DEPENDNT
X:
*ZRESID
Continue
Cancel
Help
Standardized Residual Plots
Histogram
Normal probability plot
Produce all partial plots



Unstandardized : Bağımlı değişken için modelin tahmin ettiği değerler.

Standardized: Tahmin edilen değerin standardize edilmiş değeri

Adjusted: Düzeltilmiş tahmini değerler.

S.E. Of mean predicrions: Tahmini değerlerin standart hataları

Distances : Uç nokta analizleri

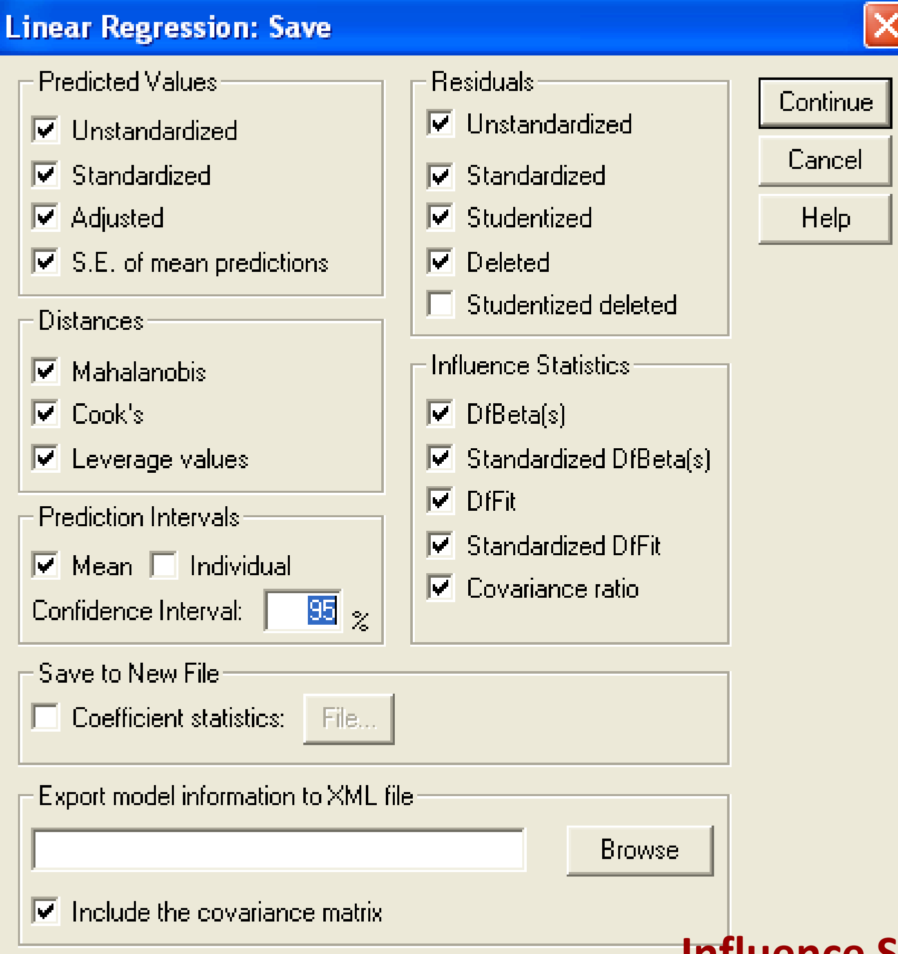
Mahalanobis: Bu değerin büyük çıkması bağımsız değişkenlerin uç değerlere sahip olduğunu gösterir.

Cook's : Bir gözlemin regresyon hesaplanmasından çıkarılması sonucu katsayıların önemli oranda değişeceğini gösterir.

Leverage values: Regresyonun uyumu üzerindeki bir noktanın etkisin ölçer.

Prediction Intervals

Mean-Individual: Ortalama tahmini-tek bir gözlem için tahmin aralığının için alt ve üst sınırları hesaplar.



Residuals (Artıklar)

Unstandardized: Gözlenen ve tahmini değerler arasındaki fark

Standardized: Stndardize artıklar (Pearson residuals ~N(0,1)).

Studentized: Student artıklar

Deleted: Bağımlı değişkenin değeri ile düzeltilmiş tahmini değer arasındaki fark.

Studentized deleted: Silinen artığın kendi standart hatasına bölümü

Influence Statistics (etki ist.)

DfBeta(s): Belli bir değişkenin çıkarılması sonucunda oluşan regresyon katsayısındaki değişim.

Stan. DfBeta(s): Herhangi bir durumun çıkarılması sonucu regresyon katsayısındaki değişim.

DfFit: Belli bir değişken çıkarıldığında tahmini değerdeki oluşan değişiklik.

StanDfFit: Herhangi bir durumun çıkarılması sonucu tahmini değerdeki değişim.

Residuals

- **Unstandardized.** The difference between an observed value and the value predicted by the model.
- **Standardized.** The residual divided by an estimate of its standard deviation. Standardized residuals, which are also known as Pearson residuals, have a mean of 0 and a standard deviation of 1.
- **Studentized.** The residual divided by an estimate of its standard deviation that varies from case to case, depending on the distance of each case's values on the independent variables from the means of the independent variables.
- **Deleted.** The residual for a case when that case is excluded from the calculation of the regression coefficients. It is the difference between the value of the dependent variable and the adjusted predicted value.
- **Studentized deleted.** The deleted residual for a case divided by its standard error. The difference between a Studentized deleted residual and its associated Studentized residual indicates how much difference eliminating a case makes on its own prediction.

Distance

- **Mahalanobis.** A measure of how much a case's values on the independent variables differ from the average of all cases. A large Mahalanobis distance identifies a case as having extreme values on one or more of the independent variables.

- **Cook's.** A measure of how much the residuals of all cases would change if a particular case were excluded from the calculation of the regression coefficients. A large Cook's D indicates that excluding a case from computation of the regression statistics changes the coefficients substantially.

- **Leverage values.** Measures the influence of a point on the fit of the regression. The centered leverage ranges from 0 (no influence on the fit) to (N-1)/N.

Influence Statistics

- **DfBeta(s).** The difference in beta value is the change in the regression coefficient that results from the exclusion of a particular case. A value is computed for each term in the model, including the constant.
- **Standardized DfBeta.** Standardized difference in beta value. The change in the regression coefficient that results from the exclusion of a particular case. You may want to examine cases with absolute values greater than 2 divided by the square root of N, where N is the number of cases. A value is computed for each term in the model, including the constant.
- **DfFit.** The difference in fit value is the change in the predicted value that results from the exclusion of a particular case.
- **Standardized DfFit.** Standardized difference in fit value. The change in the predicted value that results from the exclusion of a particular case. You may want to examine standardized values which in absolute value exceed 2 times the square root of p/N, where p is the number of parameters in the model and N is the number of cases.
- **Covariance ratio.** The ratio of the determinant of the covariance matrix with a particular case excluded from the calculation of the regression coefficients to the determinant of the covariance matrix with all cases included. If the ratio is close to 1, the case does not significantly alter the covariance matrix.

HATALARIN BAĞIMSIZLIĞI- OTOKORELASYON

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,937[a]	,878	,841	5,186	,878	23,878	3	10	,000	2,554

a. Predictors: (Constant), agirlik, yas, boy

b. Dependent Variable: kanbasýn

Hataların bağımsızlığı (otokorelasyon) herhangi bir zaman serisinin veya eşleştirilmiş zaman serilerinin değerleri arasındaki korelasyondur. Anlamlı otokorelasyon modelin yanlış yanlış tanımlandığını gösterebilir. Yani modelde önemli bir değişken unutulmuş veya fonksiyonel ilişki yanlış tanımlanmış olabilir.

Otokorelasyonun belirlenmesinde Durbin Watson (DW) istatistiği kullanılır. DW 2 civarında ise otokorelasyon yok, 0'a yakın ise pozitif otokorelasyon, 4'e yakın ise negatif otokorelasyon var demektir.

D.W.=2,554 değeri olduğundan modelde otokorelasyon yoktur.

$$DW = 2\left[1 - \frac{\sum e_t e_{t-1}}{\sum e_t^2}\right] \cong 2(1-\rho)$$

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,937[a]	,878	,841	5,186	,878	23,878	3	10	,000	2,554

a. Predictors: (Constant), agirlik, yas, boy

b. Dependent Variable: kanbasýn

Çoklu Korelasyon Katsayısı (R) : bağımsız değişkenler ile bağımlı değişken arasındaki ne derece ilişki olduğunu gösteren ölçüye denir. Gözlenen y_i değerleri ile tahmin edilen y_i değerleri arasındaki korelasyon katsayısıdır. R=0,937 oldukça yüksek bir korelasyon vardır.

Çoklu Belirtme Katsayısı (R^2) : Bağımsız değişkenlerin bağımlı değişkeni açıklama yüzdesidir (R^2=0.878). Bütün gözlemler regresyon doğrusu üzerinde yer alırsa R^2=1, eğer bağımlı ve bağımsız değişken arasında hiç doğrusal ilişki yoksa R^2=0 olur. R^2 modelin uyum iyiliği ölçütü olup, R^2=0 olması değişkenler arasında ilişki olmaması anlamına gelmez. Yani değişkenler arasında doğrusal ilişki olmadığını gösterir.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	,937[a]	,878	,841	5,186	,878	23,878	3	10	,000	2,554

a. Predictors: (Constant), agirlik, yas, boy

b. Dependent Variable: kanbasýn

Düzeltilmiş R^2 : Regresyon denklemine ilgisi olmayan bağımsız değişkenler eklendiğinde, çoklu belirtme katsayısında düzeltme yapılır. Örneklem R^2 modelin anakütleye ne kadar uyduğu ile ilgilidir. Bunun için düzeltilmiş R^2 kullanılarak modelin anakütleye uyum iyiliği daha iyi yansıtılır. Düzeltilmiş R^2 açıklayıcı değişken sayısı çok olduğunda kullanılmalıdır.

$$R^2_{düz} = 1 - \frac{\hat{\sigma}^2}{s_Y^2} = 1 - \frac{5{,}186^2}{168{,}9} = 0{,}841$$

$$R^2_{düz} = 1 - (1 - R^2)\frac{n-1}{n-k} = 1 - (1 - 0{,}878)\frac{14-1}{14-4} = 0{,}841$$

k:Toplam parametre sayısı

Std.Error of the Estimate : Artıkların standart hatası (regresyon denkleminin s.h.)

$$\hat{\sigma} = 5{,}186$$

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1926,744	3	642,248	23,878	,000[a]
	Residual	268,970	10	26,897		
	Total	2195,714	13			

a. Predictors: (Constant), agirlik, yas, boy

b. Dependent Variable: kanbasýn

ANOVA tablosu modelin bir bütün olarak anlamlı olup olmadığını test eder. P<0,05 ise kurulan regresyon modeli anlamlıdır.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-58,159	72,360		-,804	,440	-219,387	103,068					
	yas	,357	,127	,316	2,814	,018	,074	,639	,428	,665	,311	,971	1,030
	boy	,687	,539	,318	1,274	,232	-,515	1,889	,836	,374	,141	,197	5,076
	agirlik	,655	,302	,544	2,169	,055	-,018	1,327	,876	,566	,240	,194	5,143

a. Dependent Variable: kanbasın

Regresyon modeli: **Y= =-58,159+0,357Yaş+0,687Boy+0,655Ağırlık**

Standardized Coefficients-Beta: Beta bağımsız değişkenlerin önem sırasını gösterir. İşarete bakmadan en yüksek Beta değerine sahip olan değişken en önemli bağımsız değişkendir.

Eğer Xj değişkeni bir birim değil de 1 standart sapma değişseydi Y ne kadar değişir sorusuna cevap vermek için standardize Beta katsayılarına bakarak cevap vermek gerekir. Bunun için regresyondaki tüm değişkenler standardize edilip regresyon denklemi tahmin edilmelidir. SPSS'de değişkenler Analyze-Descriptive menüsünden standardize edilebilir.

Standardize edilirken ortalamadan sapmalar alındığı için standardize katsayılarda sabit (β_0) yoktur.

Değişkenleri standardize etmenin nedeni farklı biçimlerde ölçülen değişkenleri yorumlarken kolaylık sağlamaktır. Böylece X değişkeninin Y üzerindeki etkisi değişkenlerin orijinal ölçü birimleriyle değil standart sapma türünden yorumlanır. Bu kısmi etkilerin ölçü birimlerinden bağımsız olarak görülmesini ve karşılaştırılmasını salar.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-58,159	72,360		-,804	,440	-219,387	103,068					
	yas	,357	,127	,316	2,814	,018	,074	,639	,428	,665	,311	,971	1,030
	boy	,687	,539	,318	1,274	,232	-,515	1,889	,836	,374	,141	,197	5,076
	agirlik	,655	,302	,544	2,169	,055	-,018	1,327	,876	,566	,240	,194	5,143

a. Dependent Variable: kanbasın

Standardized Coefficients-Beta:

Standardize katsayılar aşağıdaki gibi hesaplanır.

$$\beta_1' = \beta_1 \frac{S_{X_1}}{S_{Y_1}}$$

$$\beta_1' = 0.357 \frac{11.512}{12.996} = 0.316$$

$$\beta_2' = 0.687 \frac{6.008}{12.996} = 0.318$$

Descriptive Statistics

	Mean	Std. Deviation	N
kanbasın	126,14	12,996	14
yas	51,71	11,512	14
boy	169,36	6,008	14
agirlik	75,6071	10,80488	14

$$\beta_3' = 0.655 \frac{10.805}{12.996} = 0.544$$

Eğer yaş (X1) bir standart sapma artarsa kan basıncı (Y) 0.316 standart sapma kadar değişecektir.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B Lower Bound	95,0% Confidence Interval for B Upper Bound	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	-58,159	72,360		-,804	,440	-219,387	103,068					
	yas	,357	,127	,316	2,814	,018	,074	,639	,428	,665	,311	,971	1,030
	boy	,687	,539	,318	1,274	,232	-,515	1,889	,836	,374	,141	,197	5,076
	agirlik	,655	,302	,544	2,169	,055	-,018	1,327	,876	,566	,240	,194	5,143

a. Dependent Variable: kanbasın

Zero-order (r) : Her bir bağımsız değişkenin bağımlı değişken (Y) ile Pearson korelasyonu

Correlations

		kanbasın	yas	boy	agirlik
Pearson Correlation	kanbasın	1,000	,428	,836	,876
	yas	,428	1,000	,097	,149
	boy	,836	,097	1,000	,895
	agirlik	,876	,149	,895	1,000
Sig. (1-tailed)	kanbasın	.	,063	,000	,000
	yas	,063	.	,370	,305
	boy	,000	,370	.	,000
	agirlik	,000	,305	,000	.
N	kanbasın	14	14	14	14
	yas	14	14	14	14
	boy	14	14	14	14
	agirlik	14	14	14	14

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-58,159	72,360		-,804	,440	-219,387	103,068					
	yas	,357	,127	,316	2,814	,018	,074	,639	,428	,665	,311	,971	1,030
	boy	,687	,539	,318	1,274	,232	-,515	1,889	,836	,374	,141	,197	5,076
	agirlik	,655	,302	,544	2,169	,055	-,018	1,327	,876	,566	,240	,194	5,143

a. Dependent Variable: kanbasın

Partial (kısmı) Korelasyon : Diğer tüm bağımsız değişkenlerin Xi ve Y üzerindeki etkisini arıttıktan sonra bağımsız değişken ile Y arasındaki korelasyon katsayısıdır.

$$r_{YX_1.X_2} = \frac{r_{YX_1} - r_{YX_2} \times r_{X_1X_2}}{\sqrt{(1 - r^2_{YX_2})(1 - r^2_{X_1X_2})}}$$

$$r_{12.34} = \frac{r_{12.3} - r_{14.3} r_{24.3}}{\sqrt{1 - r^2_{14.3}}\sqrt{1 - r^2_{24.3}}}$$

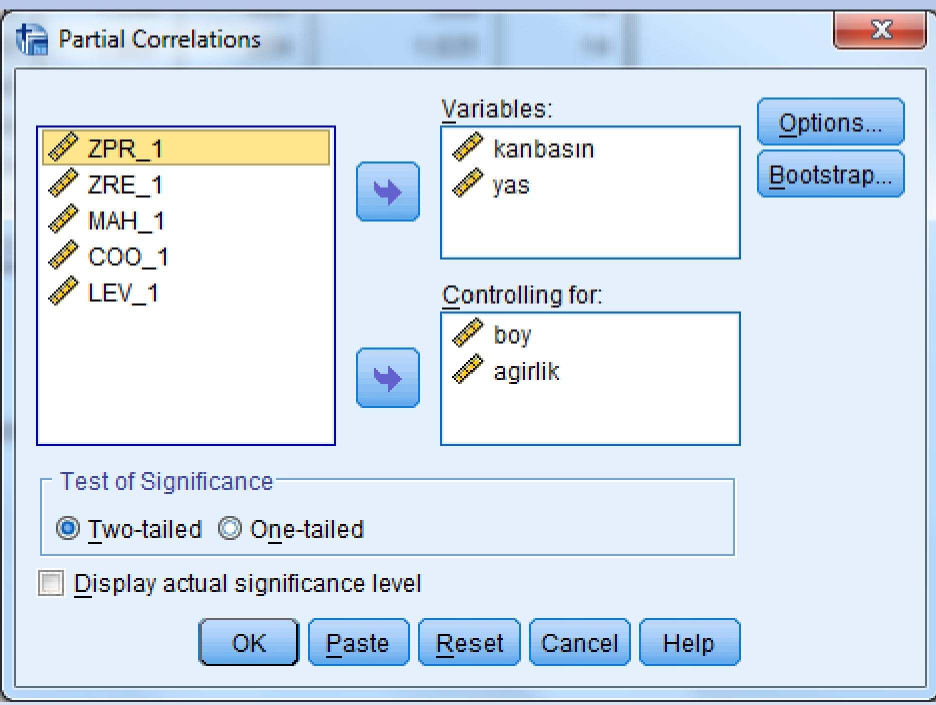
$$r^2_{12.345} = \frac{R^2_{1.2345} - R^2_{1.345}}{1 - R^2_{1.345}}$$

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B Lower Bound	95,0% Confidence Interval for B Upper Bound	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	-58,159	72,360		-,804	,440	-219,387	103,068					
	yas	,357	,127	,316	2,814	,018	,074	,639	,428	,665	,311	,971	1,030
	boy	,687	,539	,318	1,274	,232	-,515	1,889	,836	,374	,141	,197	5,076
	agirlik	,655	,302	,544	2,169	,055	-,018	1,327	,876	,566	,240	,194	5,143

a. Dependent Variable: kanbasın

Partial (kısmı) Korelasyon : Diğer tüm bağımsız değişkenlerin Xi ve Y üzerindeki etkisini arıttıktan sonra bağımsız değişken ile Y arasındaki korelasyon katsayısıdır.



Correlations

Control Variables			kanbasın	yas
boy & agirlik	kanbasın	Correlation	1,000	,665*
	yas	Correlation	,665*	1,000

*. Correlation is significant at 0.05 level

Correlations

Control Variables			kanbasın	agirlik
boy & yas	kanbasın	Correlation	1,000	,566
	agirlik	Correlation	,566	1,000

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-58,159	72,360		-,804	,440	-219,387	103,068					
	yas	,357	,127	,316	2,814	,018	,074	,639	,428	,665	,311	,971	1,030
	boy	,687	,539	,318	1,274	,232	-,515	1,889	,836	,374	,141	,197	5,076
	agirlik	,655	,302	,544	2,169	,055	-,018	1,327	,876	,566	,240	,194	5,143

a. Dependent Variable: kanbasın

Part (yarı kısmı kor.) **: Diğer tüm bağımsız değişkenlerin Xi değişkeni üzerindeki etkisini arıttıktan sonra Y ile arasındaki korelasyon katsayısıdır.**

$$r_{Y(X_1.X_2)} = \frac{r_{YX_1} - r_{YX_2} \times r_{X_1X_2}}{\sqrt{1 - r^2_{X_1X_2}}}$$

$$R^2_{1(23.4)} = R^2_{1.234} - R^2_{1.4}$$

$$R^2_{1(23.45)} = R^2_{1.2345} - R^2_{1.45}$$

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-58,159	72,360		-,804	,440	-219,387	103,068					
	yas	,357	,127	,316	2,814	,018	,074	,639	,428	,665	,311	,971	1,030
	boy	,687	,539	,318	1,274	,232	-,515	1,889	,836	,374	,141	,197	5,076
	agirlik	,655	,302	,544	2,169	,055	-,018	1,327	,876	,566	,240	,194	5,143

a. Dependent Variable: kanbasın

Bağımsız değişkenlerin bazılarının veya tümünün kendi aralarında doğrusal ilişki içinde olması durumuna **çoklu bağlantı (multicollinarity)** denir. Bu ifade doğrusal regresyon modelinden bir sapmayı ifade eder. Bu durumda parametreler için bulunan güven aralıklarının (confidence interval) büyük olmasından dolayı, o parametreler sıfırdan farklı değildir. Yani açıklayıcı değişken önemsiz demektir.

Varyans enflasyon faktörü (VIF) arttıkça regresyon katsayılarının varyansı artar. Küçük Tolerance değeri ve büyük **VIF değeri varsa (VIF>5), bağımsız değişkenler arasında çoklu bağlantı olduğunu gösterir.** Bu durumda değişkenlerden sadece birisinin modele dahil edilmesi gerekmektedir. Tabloda boy ve ağırlığın VIF değerleri yüksek, tolerance değerleri küçük, aynı zamanda iki değişken arasında yüksek korelasyon vardır. Bu durumda boy yada ağırlık değişkenlerinden biri modelden çıkarılabilir.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-58,159	72,360		-,804	,440	-219,387	103,068					
	yas	,357	,127	,316	2,814	,018	,074	,639	,428	,665	,311	,971	1,030
	boy	,687	,539	,318	1,274	,232	-,515	1,889	,836	,374	,141	,197	5,076
	agirlik	,655	,302	,544	2,169	,055	-,018	1,327	,876	,566	,240	,194	5,143

a. Dependent Variable: kanbasın

VIF(Yaş) : Yas bağımlı değişken alınır ve diğer bağımsız değişkenlerle (boy, ağırlık) çoklu korelasyon katsayısı (R_1^2) hesaplanır.

$$VIF(yas)=1/ (1-R_1^2)$$

VIF(boy) : Boy bağımlı değişken alınır ve diğer bağımsız değişkenler (yas, ağırlık) arasındaki çoklu korelasyon katsayısı (R_2^2) hesaplanır.

$$VIF(boy)=1/(1-R_2^2)$$

VIF(ağırlık) : Ağırlık bağımlı değişken alınır ve diğer bağımsız değişkenler (yas, boy) arasındaki çoklu korelasyon katsayısı (R^2) hesaplanır.

$$VIF(ağırlık)=1/(1-R_3^2)$$

Bağımlı değişken ile bağımsız değişken arasında ilişki yoksa R^2=0 ve VIF=1 olur. R^2=0,9 ise VIF=1/(1-0,9)=10 olur (Webster, 1995, 683-684).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-58,159	72,360		-,804	,440	-219,387	103,068					
	yas	,357	,127	,316	2,814	,018	,074	,639	,428	,665	,311	,971	1,030
	boy	,687	,539	,318	1,274	,232	-,515	1,889	,836	,374	,141	,197	5,076
	agirlik	,655	,302	,544	2,169	,055	-,018	1,327	,876	,566	,240	,194	5,143

a. Dependent Variable: kanbasın

VIF(Yaş) : Yas bağımlı değişken ve boy ile ağırlık bağımsız değişken alınırsa R_1^2=0,152 bulunur.

VIF(yas)=1/ (1-0,029)= 1,030

VIF(boy) : Boy bağımlı değişken alınır ve yas ile ağırlık bağımsız değişken alınırsa R_2^2=0,803 hesaplanır.

VIF(boy)=1/(1-0,803)=5,076

VIF(ağırlık) : Ağırlık bağımlı değişken alınır ve diğer yas ile boy bağımsız değişken alınırsa R_3^2=0,806 bulunur.

VIF(ağırlık)=1/(1-0,806)=5,143

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-58,159	72,360		-,804	,440	-219,387	103,068					
	yas	,357	,127	,316	2,814	,018	,074	,639	,428	,665	,311	,971	1,030
	boy	,687	,539	,318	1,274	,232	-,515	1,889	,836	,374	,141	,197	5,076
	agirlik	,655	,302	,544	2,169	,055	-,018	1,327	,876	,566	,240	,194	5,143

a. Dependent Variable: kanbasın

Çoklu bağlantı problemінin hesaplanmasında kullanılan bit diğer yöntem ise değişkenlerin Toleranslarını (Tolerance) hesaplamaktır.

$$\textbf{Tolerance (T)}=1-R_i^2$$

Küçük tolerans değeri büyük VIF değeri demektir (Gujatari, 1995, 338-339).

Coefficient Correlations[a]

Model			agirlik	yas	boy
1	Correlations	agirlik	1,000	-,140	-,895
		yas	-,140	1,000	,082
		boy	-,895	,082	1,000
	Covariances	agirlik	,091	-,005	-,146
		yas	-,005	,016	,006
		boy	-,146	,006	,291

a. Dependent Variable: kanbasýn

Boy ile ağırlık arasında yüksek korelasyon vardır. Bu durum çoklu bağlantı problemi için bir ip ucudur.

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	yas	boy	agirlik
1	1	3,956	1,000	,00	,00	,00	,00
	2	,033	10,912	,00	,96	,00	,01
	3	,010	19,749	,01	,02	,00	,21
	4	,000	169,311	,99	,01	1,00	,77

a. Dependent Variable: kanbasýn

Çoklu bağlantı özdeğer, koşul endeksi (Condition Index) ve değişim oranları (Variance Proportions) ile de bulunabilir. Ağırlık değişkeninin %1'lik değişimi 2. özdeğerle, %21'lik değişimi 3. özdeğerle ve %77'lik değişimi 4. özdeğerle ilişkilidir. Aynı özdeğerde yüksek oranı olan değişkenlere bakılır. 4.özdeğer sabitteki varyansın %99'unu, boydaki değişimin %100'ünü, ağırlıktaki değişimin ise %77'sini açıklar. Bu bize değişkenlerin birbiri ile bağımlı olduğunu gösterir. 4.özdeğerin en küçük olması ve durum endeksinin ise en yüksek olması önemlidir.

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	yas	boy	agirlik
1	1	3,956	1,000	,00	,00	,00	,00
	2	,033	10,912	,00	,96	,00	,01
	3	,010	19,749	,01	,02	,00	,21
	4	,000	169,311	,99	,01	1,00	,77

a. Dependent Variable: kanbasýn

Çoklu doğrusal bağlantı probleminin bulunmasında kullanılan bir diğer yöntem ise koşul endeksidir (Condition Index:CI).

$$CI = \sqrt{V_{max} / V_{x_i}}$$

V_{max} : Maksimum açıklanan varyans (maksimum özdeğer)
V_{xi} : i.değişken tarafından açıklanan toplam varyans (x_i değişkeninin özdeğeri)

CI değeri 10-30 arasında ise orta düzeyde, 30'dan büyükse yüksek çoklu bağlantı problemi vardır.

ÇOKLU BAĞLANTI (Collinearity Diagnostics)

Bağımsız değişkenler arasındaki yakın doğrusal bağımlılık (çoklu bağlantı) regresyon modellerini etkiyen önemli bir problemdir. Çoklu bağlantı regresyon katsayılarının doğruluğunu etkiler.

VIF (Variance Inflation Factor) : VIF değeri 5'den büyükse (bazı kaynaklarda >10) çoklu bağlantı problemi olduğu belirtilir (Montgomery, Peck, Vininig, 2012).

Çoklu Doğrusal Bağlantı Probleminin Çözümü

1. **Adımsal regresyon yöntemi ile bir veya birkaç değişken modelden çıkarılır.**
2. **Örnek sayısı artırılabilir.**
3. **Birbiri ile ilişkili olan iki değişkenin toplamı alınarak modele dahil edilir.**
4. **Değişkenler farkları alınarak dönüştürülebilir.**

(Gajarati, 1995, 339-334).

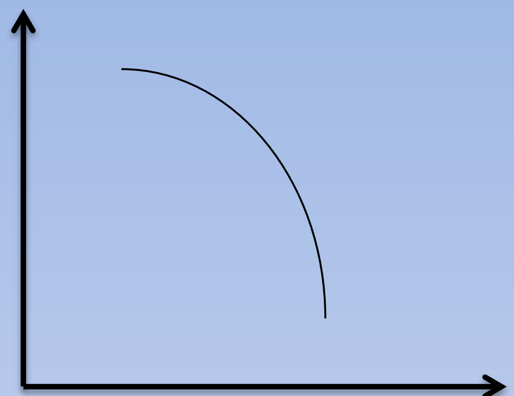
Polinomlu regresyon modellerinde çoklu bağlantı problemi değişkenlerin ortalamadan farkları alınarak dönüşüm yapılmaktadır.

Doğrusallık

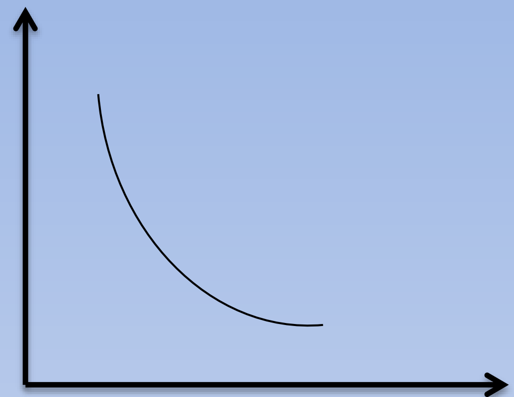
Doğrusallık, iki değişken arasında bir doğru ile özetlenebilen bir ilişki olduğu anlamına gelir. Doğrusallık çok değişkenli analizler açısından önemlidir.

Regresyon analizindeki artık grafiği (residuals plot) doğrusal olmayan ilişkileri de gözlemeye yarar. Artıklar çok değişkenli analizlerce açıklanamayan değerlerdir. Artıklar bir değişkene ilişkin elde edilen değerler ile tahmin değerleri arasındaki farktır. Standardize edilmiş artık grafikleri ile doğrusallık kontrol edilebilir. Grafikte artıklar bazı tahmin değerleri için çizginin altında bazıları için de çizginin üstünde yer alıyorsa, doğrusallık varsayımı sağlanmıyor demektir (Tabachnick ve Fidell, 1996). *Eğer noktalar sıfır çizgisi etrafında kümeleniyorsa, doğrusallık varsayımı sağlanır.*

Doğrusallık Varsayımı İçin Uygun Dönüşümler

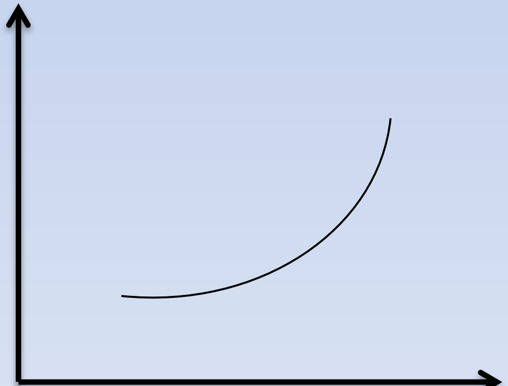


a) Y^2 , X^2



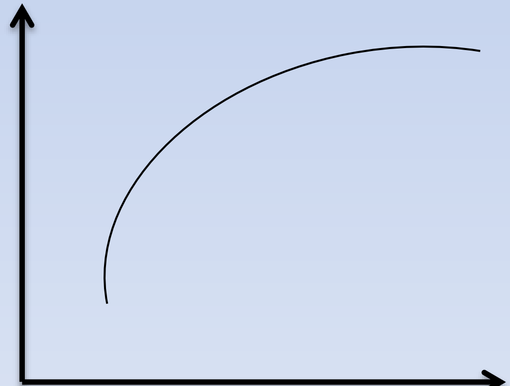
b) $\log(Y)$, $\log(X)$

$\sqrt{Y}$, $\sqrt{X}$



c) $\log(Y)$, X^2

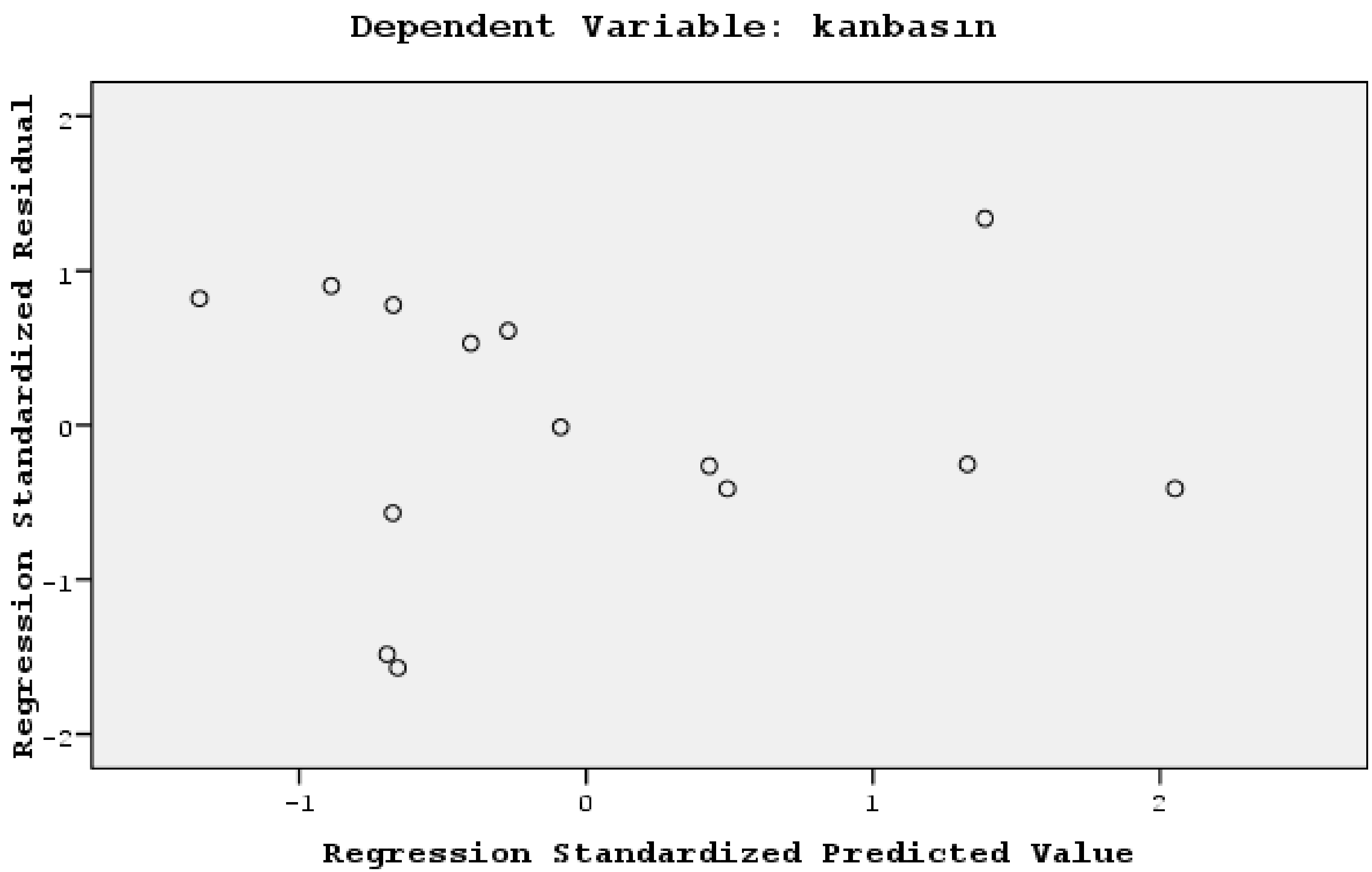
$\sqrt{Y}$, X^2



d) Y^2 , $\log(X)$

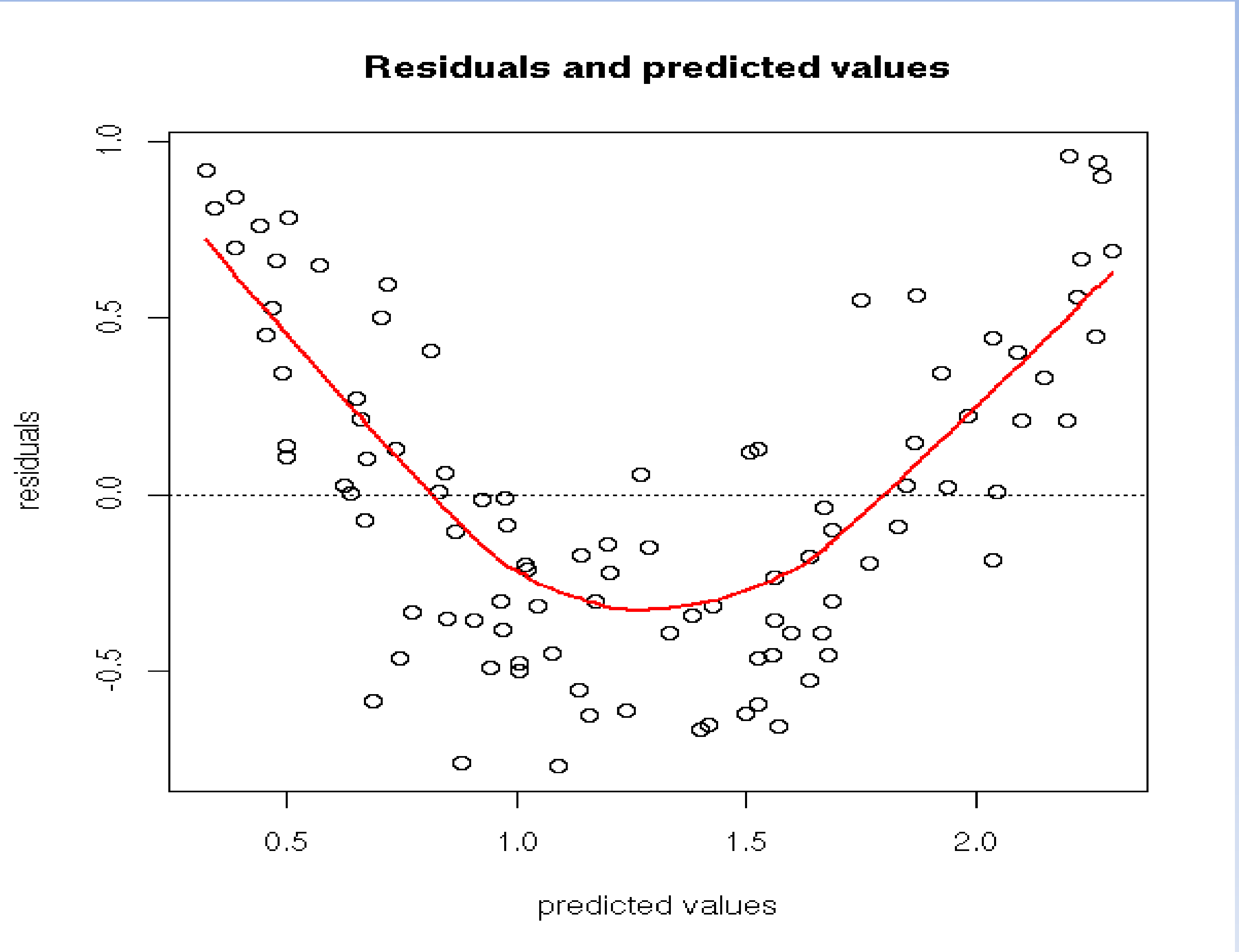
Y^2 , $\sqrt{X}$

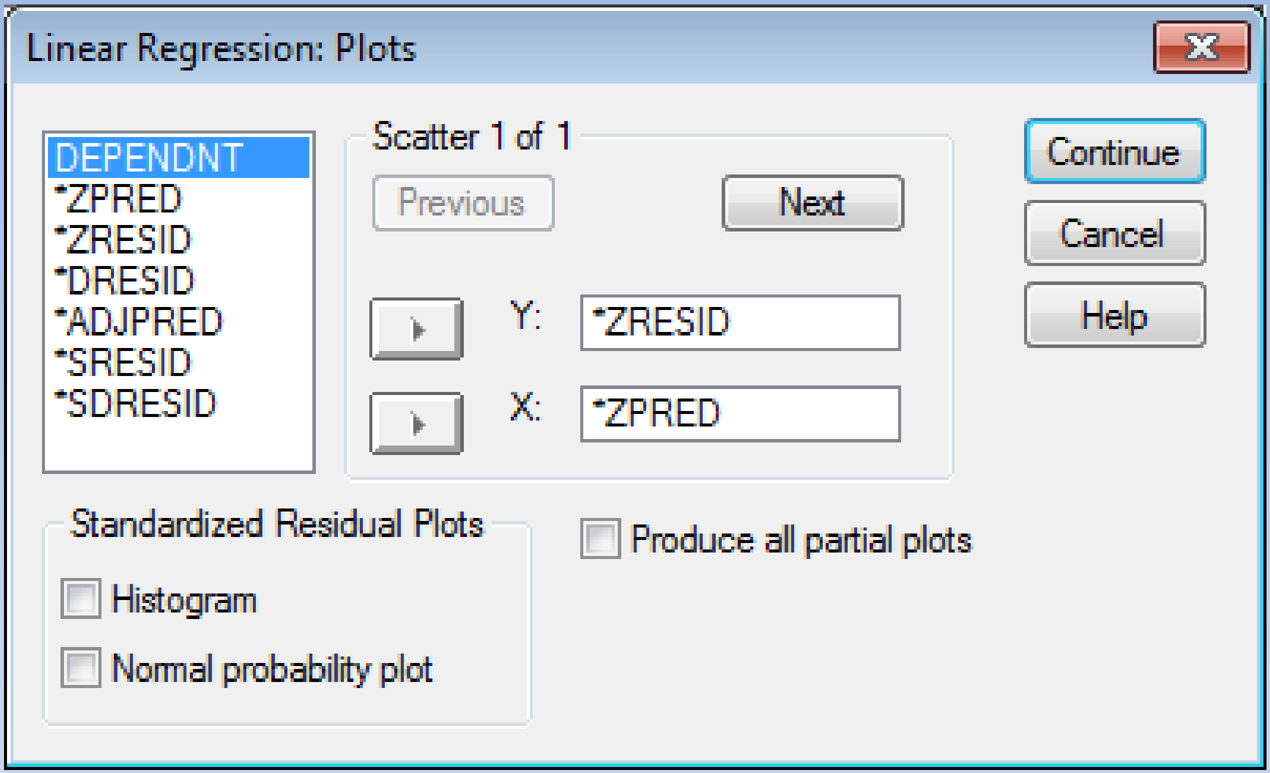
Doğrusallık



Hata terimleri ve tahmin edilen değerler arasındaki serpme grafiğinde, noktalar rasgele ve bir yığılma olmadan dağılmalıdır. Eğer noktalar bir eğri gösteriyorsa doğrusal ilişki varsayımı sağlanmaz. Serpme grafiğinde belirli bir görünüm oluşmuş ise doğrusal olmayan bir ilişki var demektir.

Non Linear (Doğrusallık sağlanmıyor)




Linear Regression: Plots
DEPENDNT
*ZPRED
*ZRESID
*DRESID
*ADJPRED
*SRESID
*SDRESID
Scatter 1 of 1
Previous
Next
Y:
*ZRESID
X:
*ZPRED
Continue
Cancel
Help
Standardized Residual Plots
Histogram
Normal probability plot
Produce all partial plots

Sabit varyans varsayımı: X değeri ne olursa olsun hata teriminin varyansı hep aynıdır. Yani X ile Y arasındaki ortala ilişki doğrusu (regresyon doğrusu) dolayındaki oynaklık bütün X değerleri boyunca aynıdır. X değerlerine göre ne artar, nede azalır.

Değişen varyans durumunda EKK tahminleri yansız olmakta, ancak varyans ve kovaryans tahminleri etkin (minimum varyanslı) olmadığından istatistik hipotez testleri geçerliliğini kaybetmektedir. Ayrıca belirli bir anlamlılık düzeyindeki tahmin ve öngörü aralıkları genişlemektedir. Değişen varyans durumunda EKK yerine çeşitli yaklaşımlar kullanılır (örneğin ağırlıklı en küçük kareler yöntemi gibi).

Değişen varyansı belirlemede çeşitli yöntemler kullanılır.

Park Testi, Glejser Testi, Harrison – McCabe Testi, Goldfeld – Quandt Testi , Breusch – Pagan – Godfrey Testi , White Testi, Bartlett Testi, Spearman Sıra Korelasyon Testi, ARCH (LM) ve GARCH

Değişen Varyans Testleri

PARAMETRİK TESTLER

EKK'ya Dayalı

- Park testi
- Glejser testi
- Golfeld-Quandt testi
- Breush-Pagan-Godfrey
- White testi
- Harrison-Mccabe testi
- Bartlett testi
- Reset testi

Maksimum Olabilirliğe Dayalı

- LR (Olabilirlik oranı) testi
- LM (Lagrange çarpanı) testi
- Wald testi

Arch-Garch testi

NON-PARAMETRİK TESTLER (Parametrik Olmayan)

Sıra Korelasyon Testi

Peak Testi

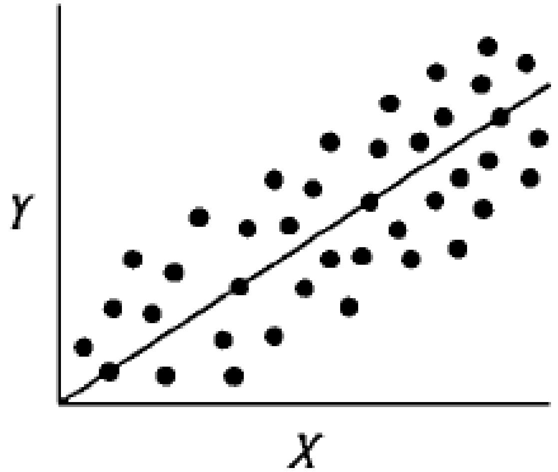
1. **Logaritmik Dönüşüm ($Y^*=LnY$):** Bu dönüşüm negatif sayıların logaritması alınmayacağı için sadece pozitif birim değerli değişkenlere uygulanabilmektedir. y artarken y'nin varyansı da artıyorsa varyansı durağanlaştırmak, y'nin hataları sağa çarpıksa y'nin dağılımını normalleştirmek ve bağımlı değişken ile bağımsız değişken arasında sürekli artan bir eğim söz konusu olması durumunda ise modeli doğrusallaştırmak için kullanılmaktadır.
2. **Karekök Dönüşümü ($Y^*=Y^{0.5}$):** Varyans y'nin ortalaması ile orantılı ise varyansı durağanlaştırmak için kullanılmaktadır. Bu dönüşüm özellikle bağımlı değişkenin Poisson dağılımına uyması durumunda kullanılmaktadır.
3. **Hiperbolik Dönüşüm($Y^*=1/Y$):** Varyans y'nin dördüncü dereceden kuvvetiyle orantılı ise (y'nin ilk değerleri arasında aşırı farklar varsa) varyansı durağanlaştırmak için kullanılmaktadır. Bu dönüşümle serideki sapan değerler etkisiz hale getirilmektedir. Çünkü büyük sayıların tersi sıfıra daha yakın olacağından y'deki sapan değerler y^*'da önemsiz olmaktadır.
4. **Kare Dönüşümü ($Y^*=Y^2$):** Varyans y'nin ortalamasına göre azalıyorsa varyansı durağanlaştırmak, bağımlı değişkenin hata değerleri sola çarpıksa bağımlı değişkeni normalleştirmek ve bağımsız değişkenlerden bazıları bağımlı değişken ile aşağıya doğru eğrisel bir ilişki (bağımsız değişken artarken eğimin ani düşmesi) göstermesi durumunda modeli doğrusallaştırmak için kullanılmaktadır.
5. **Arcsin Dönüşümü ($Y^*=ArcsinY^{0.5}=Sin^{-1}.Y^{0.5}$):** y, oran veya göreceli bir büyüklük ise varyansı durağanlaştırmak için kullanılmaktadır.

Sabit varyans varsayımı sağlanmadığında yapılacak işlemler:

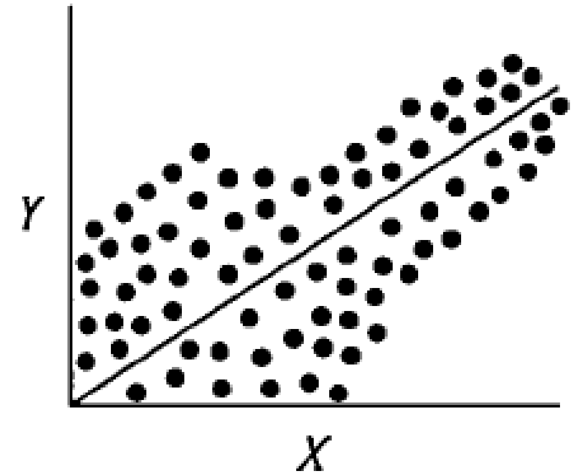
Değişen varyans probleminin çözümü için değişkenlere dönüşüm uygulanır. Değişkenlere dönüşüm uygulamanın doğrusallaştırmak, normalleştirmek ve durağanlaştırmak (sabit varyans) gibi üç temel amacı vardır

Değişen varyansın(Heteroskedasticity) nedenleri:

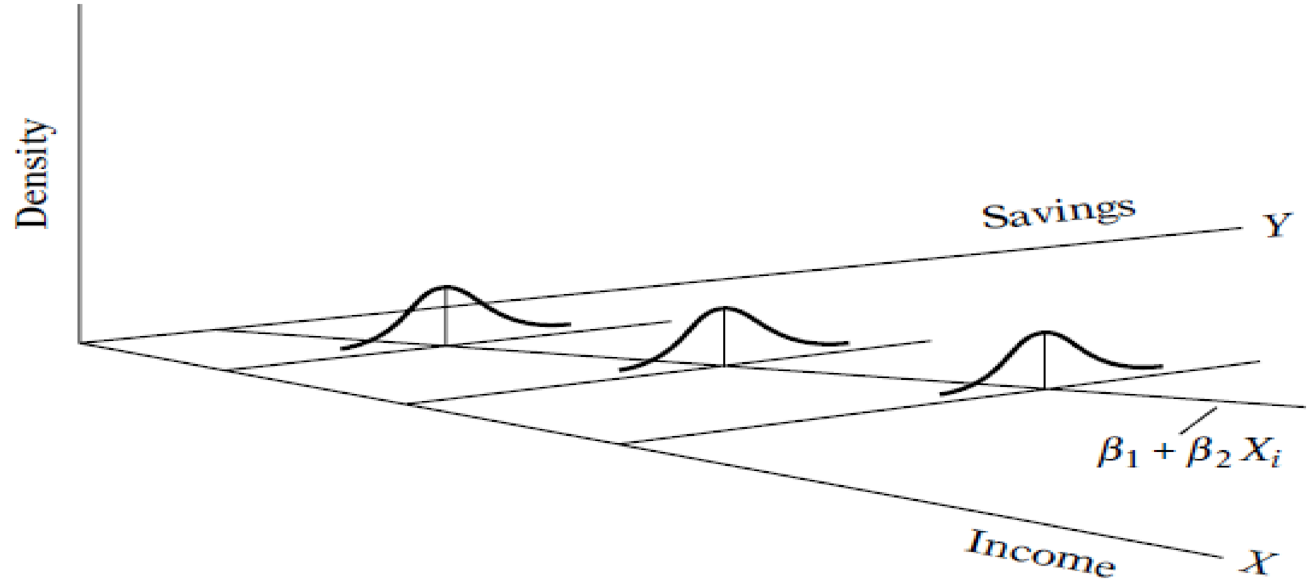
- Modele katılmış değişkenlerin çarpık olması,
- Yanlış veri dönüştürmesi,
- Yanlış fonksiyon seçimi, (doğrusal modele karşı çifte-log model)



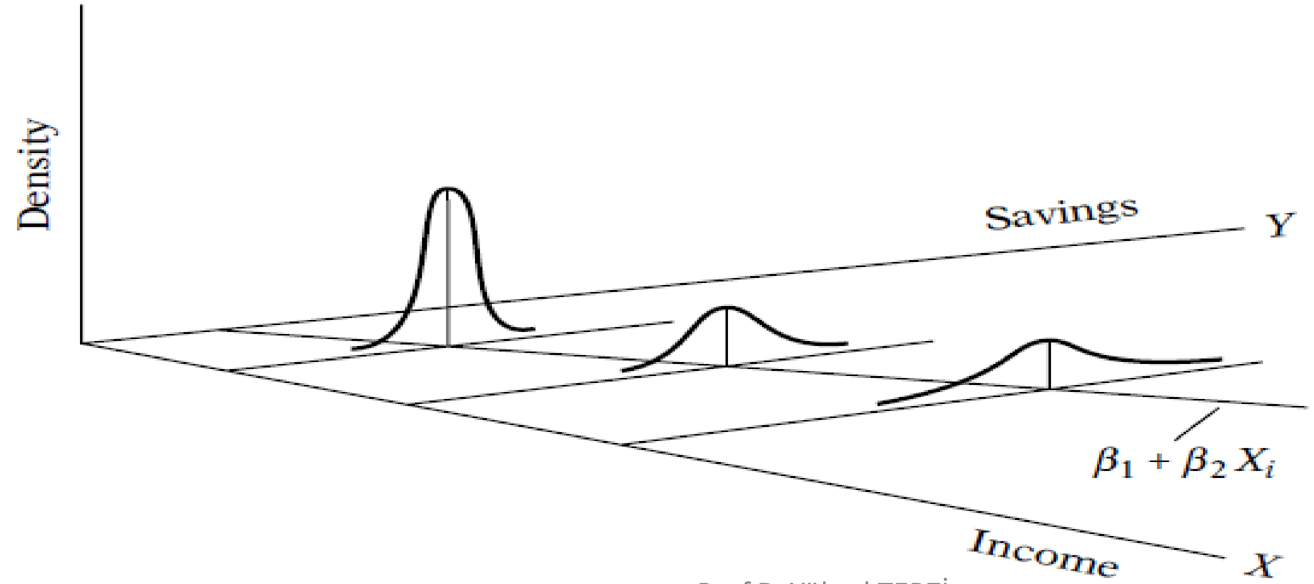
(a) Homoscedastic data



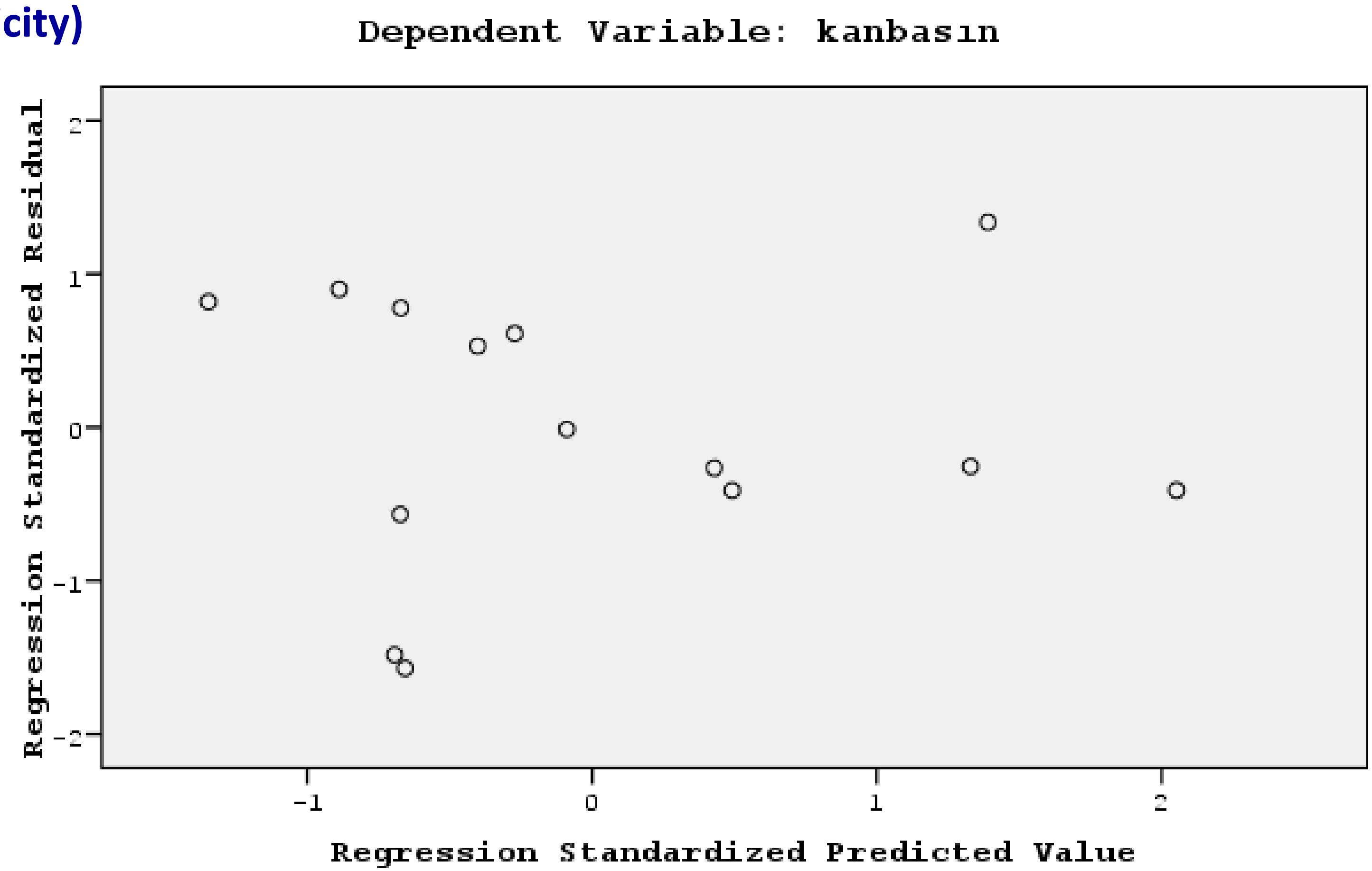
(b) Heteroscedastic data



Homoscedastic disturbances.

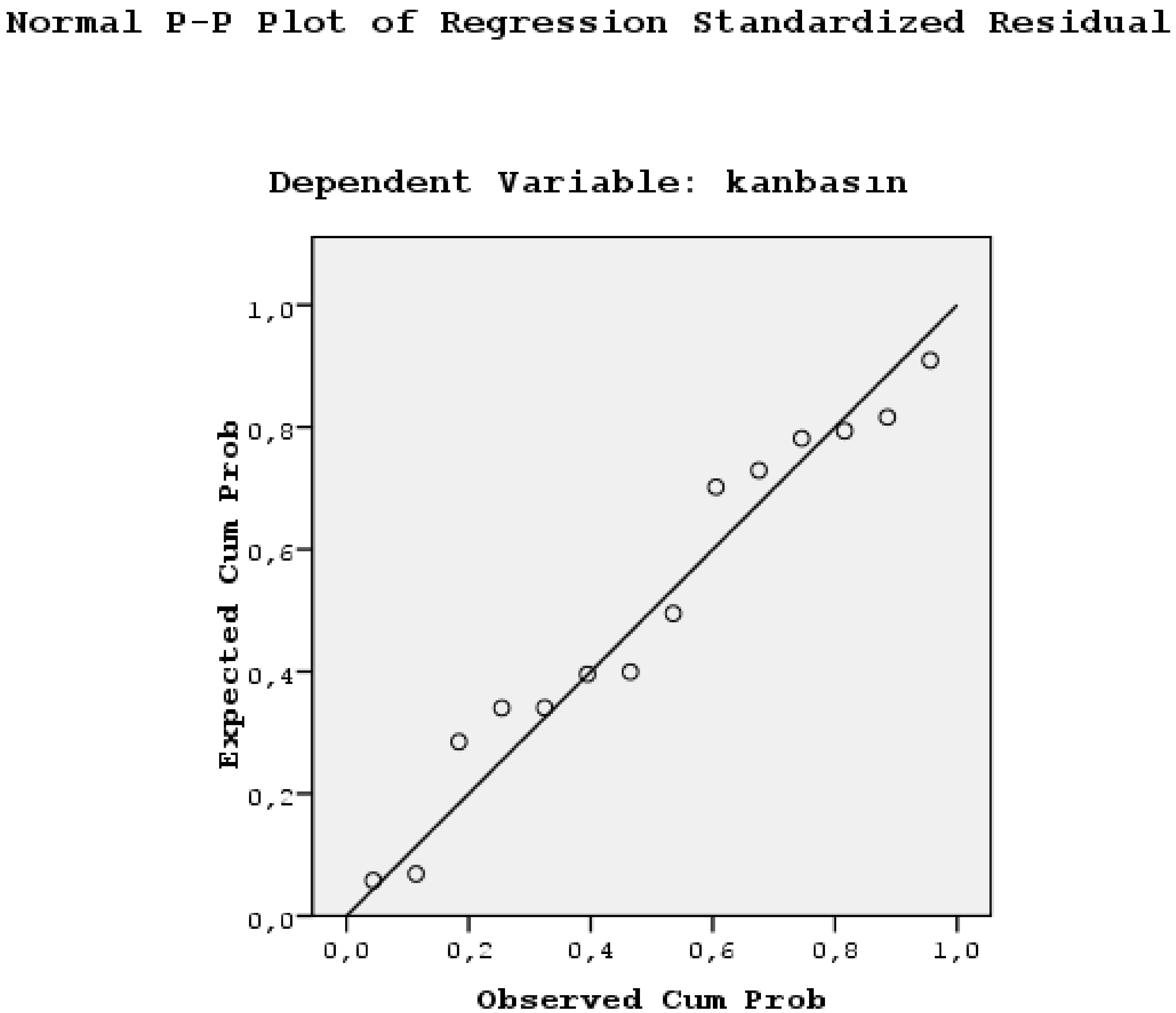


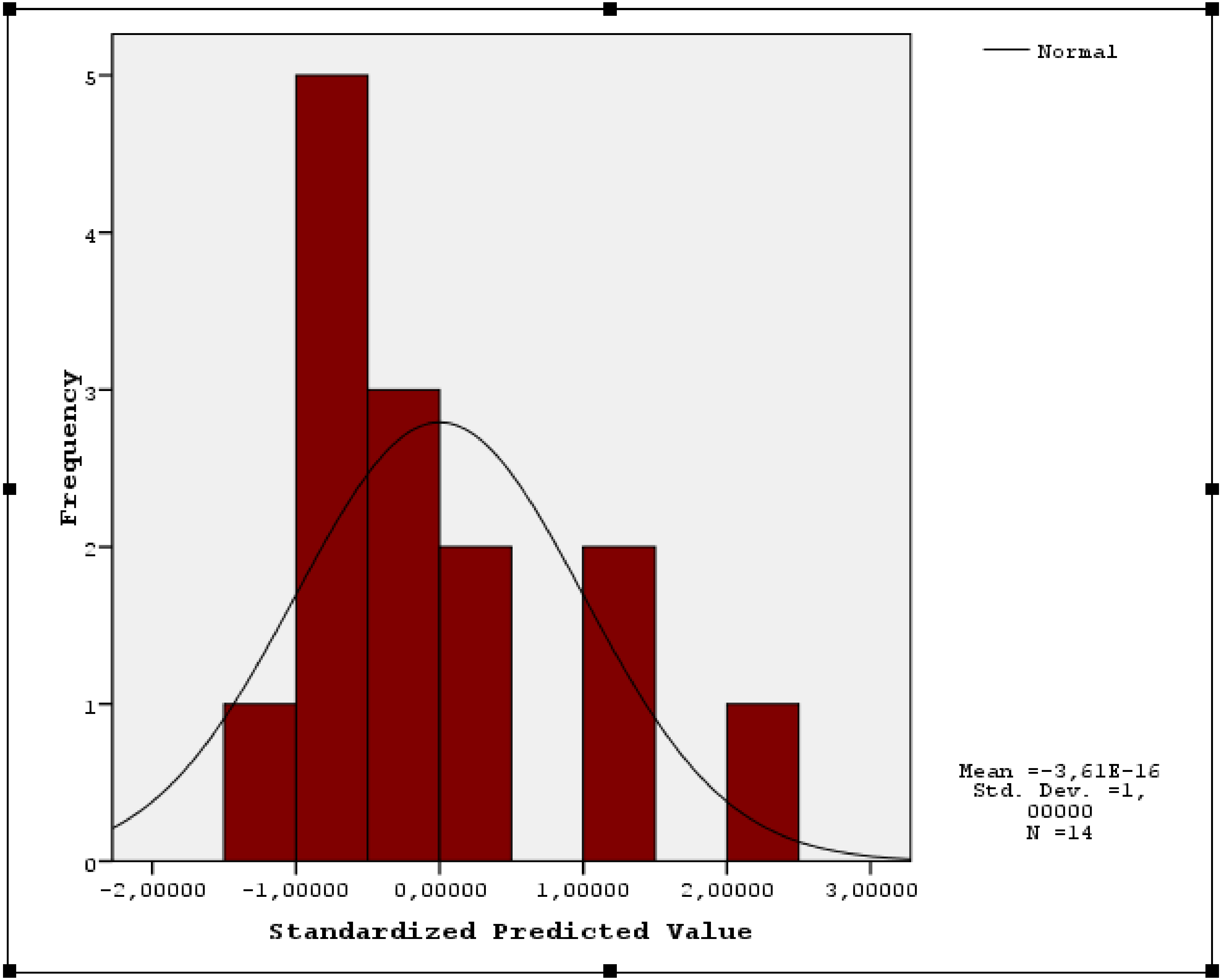
Sabit Varyans Varsayımı (Homoscedasticity)



Eğer hata terimlerinin sabit varyansı varsa, noktaların dikey yöndeki yayılması, serpme grafiğinin her bölümünde aynı olur. Şekilde noktalar rasgele dağılmış ve bir eğri veya huni şekli oluşturmadıklarından sabit varyans varsayımı sağlanmaktadır.

Normallik


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: kanbasın
Expected Cum Prob
Observed Cum Prob
0,0
0,2
0,4
0,6
0,8
1,0



Normal
Frequency
5
4
3
2
1
0
-2,00000
-1,00000
0,00000
1,00000
2,00000
3,00000
Standardized Predicted Value
Mean =-3,61E-16
Std. Dev. =1,
00000
N =14

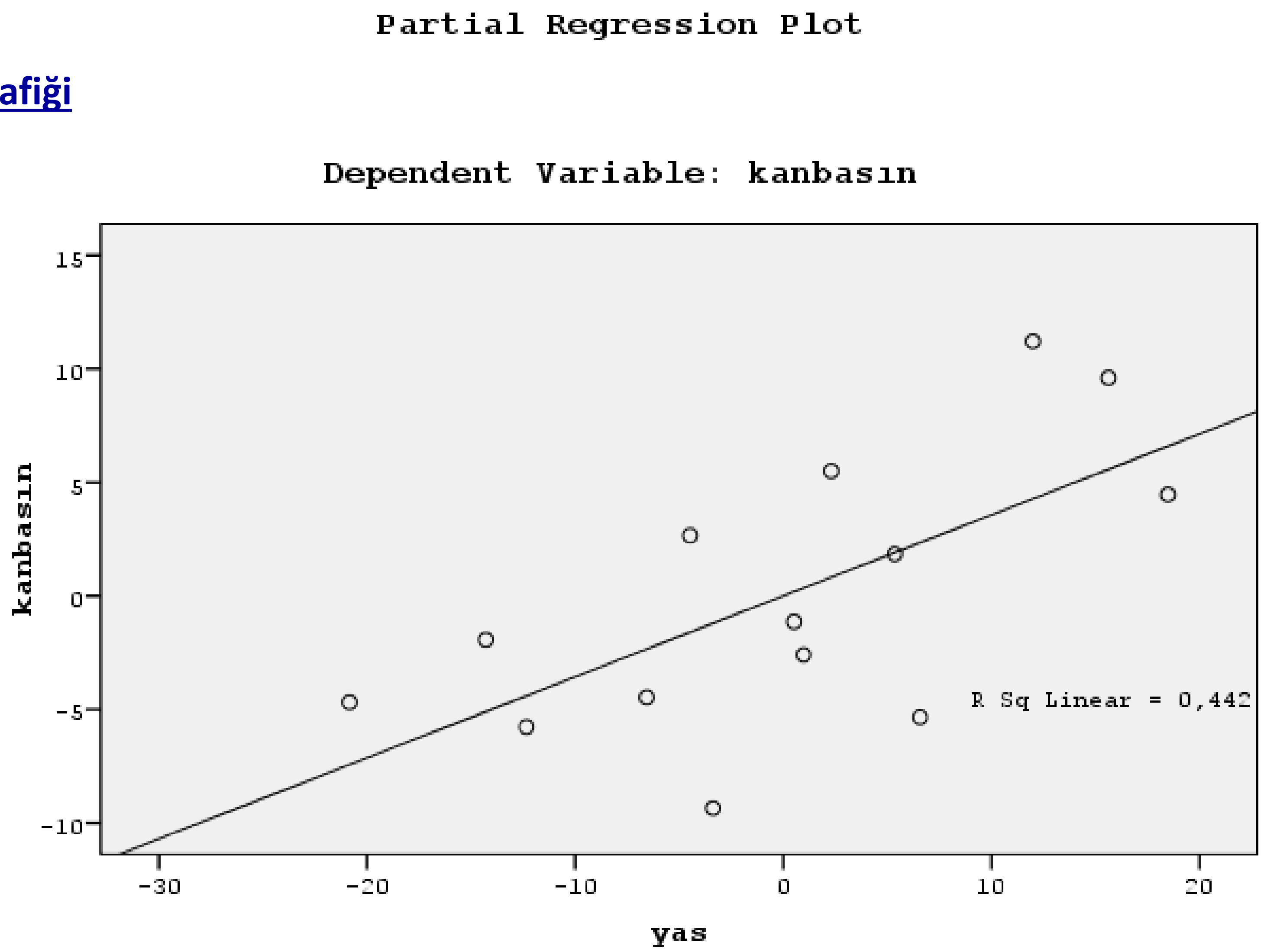
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	,156	14	,200*	,940	14	,412
Standardized Predicted Value	,179	14	,200*	,905	14	,131

*. This is a lower bound of the true significance.

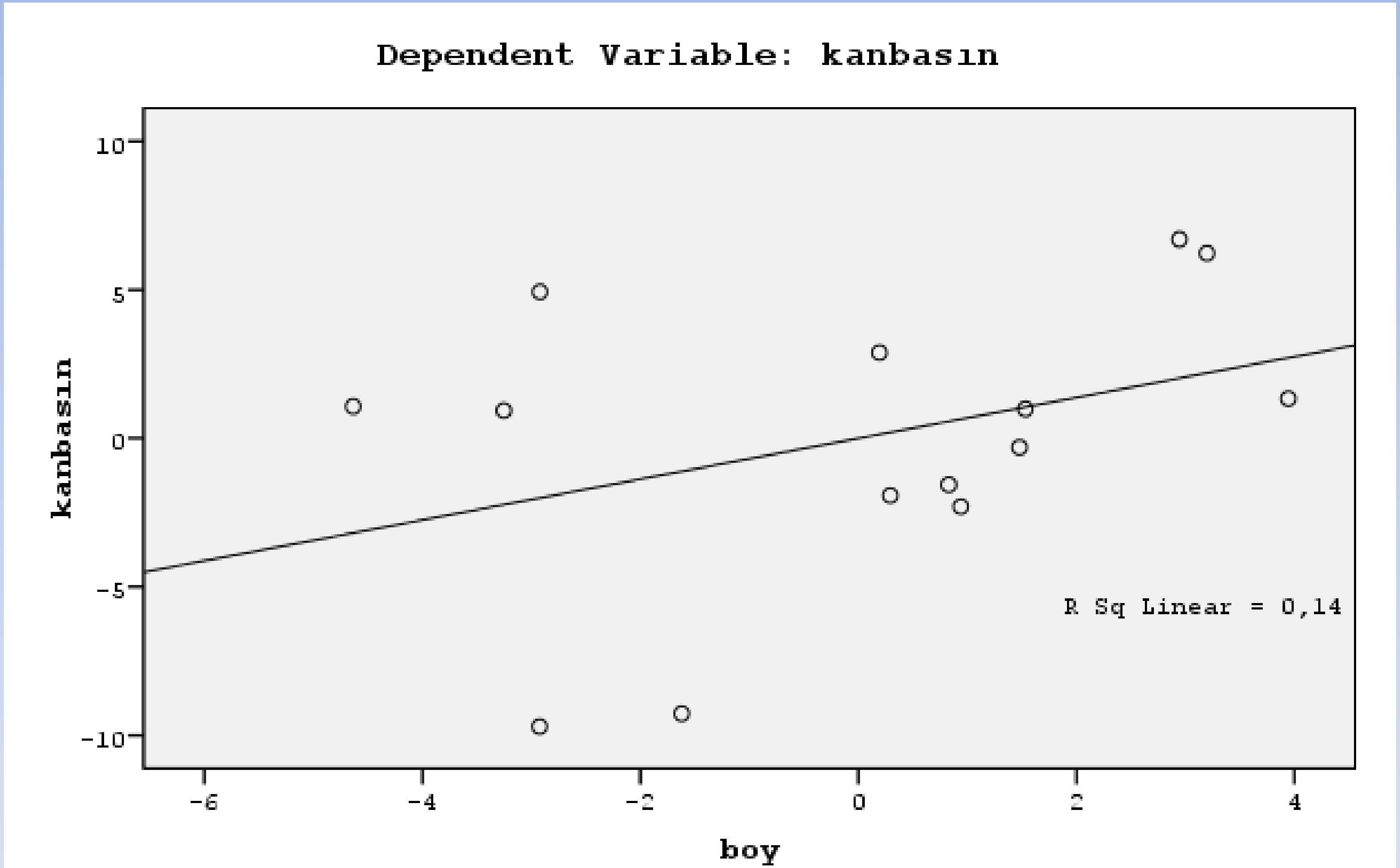
a. Lilliefors Significance Correction

Kısmi Regresyon Grafiği



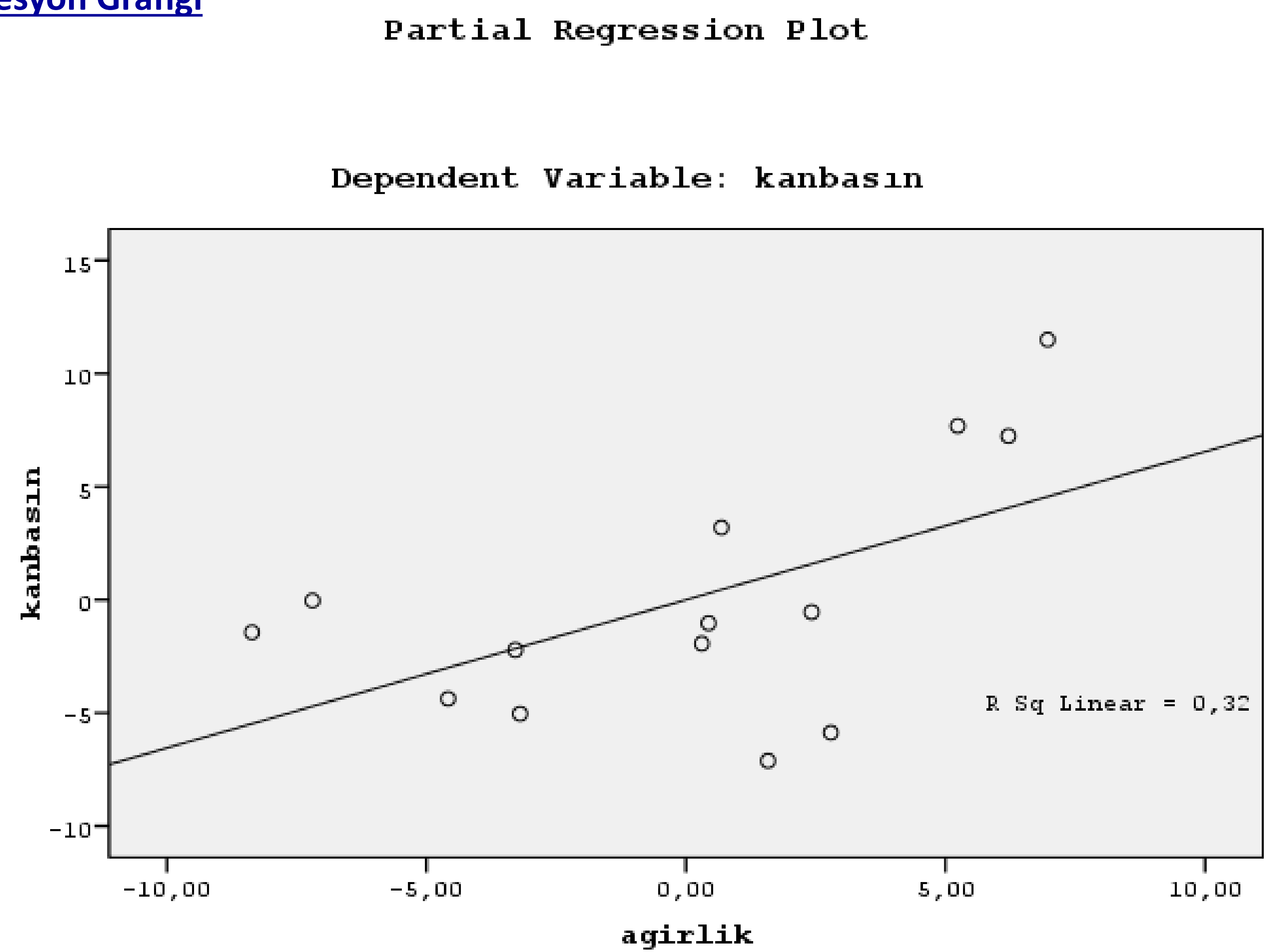
Kısmi regresyon grafiği herbir açıklayıcı değişken için çizilir ve dağılımın doğrusallıktan uzaklaşması durumunda başka açıklayıcı değişkenlerin modele ilave edilmesi gerekebilir.

Kısmi Regresyon Grafiği



Boy değişkeninde doğrusallıkta uzaklaşma görülmektedir. R^2'side oldukça düşüktür.

Kısmi Regresyon Grafiği



AYKIRI DEĞERLER (Outliers)

Hawkins(1980)'e göre bir aykırı değer diğer gözlemlerden oldukça sapan ve başka bir mekanizma tarafından oluştuğu konusunda kuşku uyandıran bir gözlemdir.

Regresyon analizinde verilerde aykırı değer varsa EKK kestiricilerin istediğimiz özelliklere sahip olmamasına neden olacaktır. Bu durumda ya klasik tanılama (diagnostics) yöntemleriyle aykırı değerlerin bulunması ve kalan gözlemler üzerinde klasik yöntemlerle kestirim yapılır yada sağlam (roboust) yöntemler kullanılarak analiz sonuçları üzerinde büyük etki sahibi olan saykırı değerlerin etkileri azaltılmaya çalışılır.

Regresyon analizinde 4 farklı aykırı değer durumu vardır.

1. Bağımlı değişken (Y) yönünde extreme gözlemler (aykırı değerler).
2. Bağımsız değişken (ler) (X) yönündeki extreme gözlemler (yüksek kaldıraç-high leverage).
3. Hem bağımlı hem de bağımsız değişkenler yönünde olan extreme gözlemler.
4. Etkili gözlemler (influential observations).

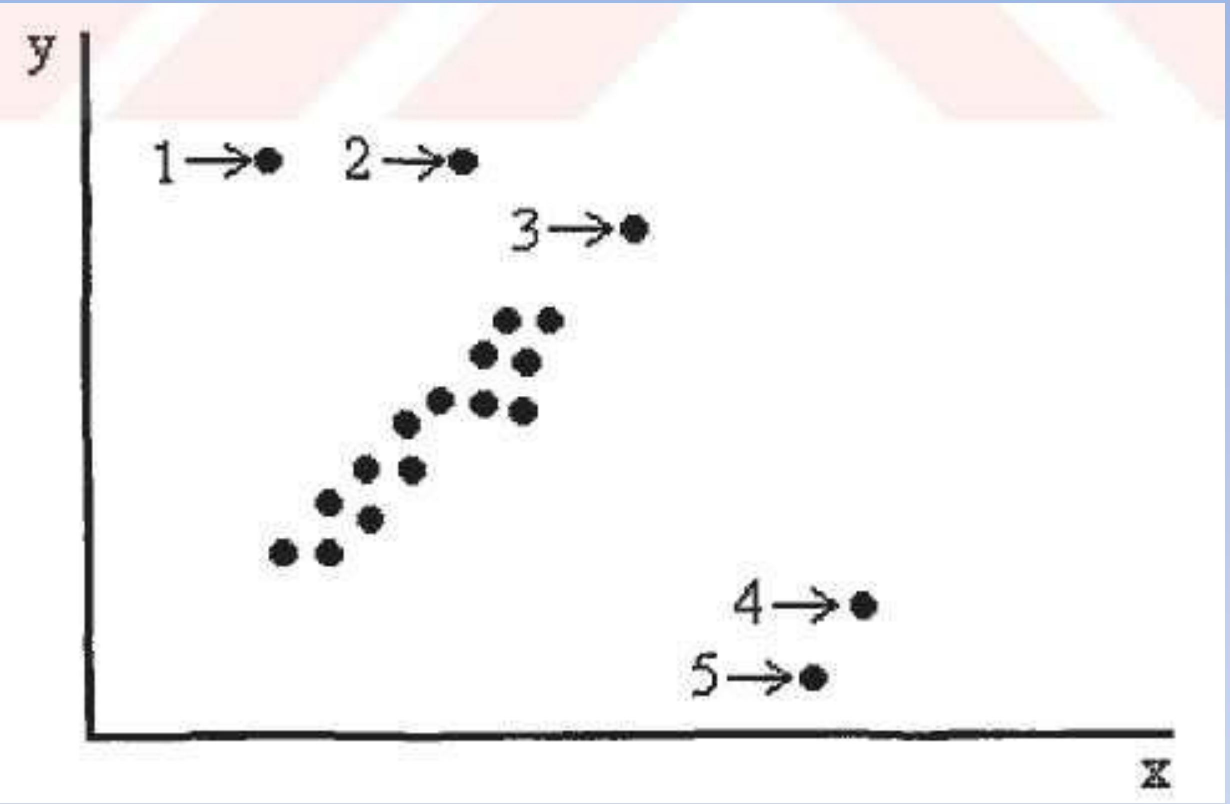
Aykırı değer ve etkili gözlem arasındaki ilişki:

- **Aykırı değerler veya yüksek kaldıraç değerleri etkili gözlem olmak zorunda değildir.**

- **Etkili gözlemler genellikle aykırı değerler veya yüksek kaldıraç değerlerdir.**

(Chatterjee and Hadi, 1988)

Aykırı Gözlemler



1 ve 2 numaralı gözlemler y yönünde aykırı değer, 3-4-5 numaralı gözlemler x yönünde kaldıraç değerlerdir (3 iyi kaldıraç, 4-5 kötü kaldıraç noktasıdır) (Neter et al, 1996).

Aykırı Değerler (Outliers)

Bir gözlem; $\overline{X} \pm 2SS$ veya $\overline{X} \pm 3SS$

aralığı dışında kalıyorsa aykırı değer olarak belirtilir. Aykırı değerler bağımlı değişken üzerinde incelenir.

Aykırı değerlerin sebepleri:

- Hatalı ölçüm
- Veri girişindeki hata
- Ölçüm yapılan araçtaki hata
- Nadir görülen olaylar

Aykırı değerlerin regresyonda sebep olduğu sorunlar:

- Örneklem ortalamasına ve varyansına bağlı istatistiksel testleri saptırır.
- Regresyon katsayılarının değerleri, t ve F değerleri, R^2 ve AKO aykırı değerlerden etkilenir.
- Kestiricilerin yanlılığına sebep olur.
- AKT'yi min yapma ilkesi etkilenir.
- İstatistiksel anlamlılıkta p değerini saptırırlar.
- Yanlış karar vermeye sebep olurlar.

(Aşıkgil, 2006).

Ölçülmüş bir değerin kaydında bir hata varsa, eğer örneklem noktası geçersizse ya da seçilen örneklem kitlesinin bir parçası değilse o zaman bu gözlemin çıkarılması uygun olabilir (Montgomery ve ark., 2012).

Aykırı gözlemlerin belirlenmesinde kullanılan bazı istatistik testler:

1. **Dixon Testi**
2. **Rosner Testi**
3. **Discordance Testi**
4. **Grubbs Testi**
5. **Wals Testi**

Testler	Normallik varsayımı	Örnek büyüklüğü	Aykırı gözlem sayısı	Test sınıfı
Dixon Test	Normal	$3 \leq n \leq 25$	Tek	Parametrik
Rosner Test	Normal	$n \geq 25$	Çok	Parametrik
Discordance Test	Normal	$n \leq 50$	Tek	Parametrik
Grubbs Test	Normal	$n \leq 50$	Tek/Çok	Parametrik
Walsh Test	Normal olmayan	$60 \leq n \leq 220 \quad \alpha = 0.10$ $n > 220 \quad \alpha = 0.05$	Çok	Parametrik olmayan

RES_1	ZRE_1
-2,94565	-,56798
4,03259	,77756
-2,13484	-,41164
3,16896	,61103
-7,69494	-1,48372
-1,32041	-,25460
2,74774	,52982
6,93521	1,33723
-8,15290	-1,57203
-1,37118	-,26439
4,67362	,90116
-2,12843	-,41040
-,06438	-,01241
4,25461	,82037

Standardize Artıklar
(Standardized Residual:ZRE)

$$d_i = \frac{e_i}{\sqrt{AKO}}$$

Standardize artıklar (ZRE) birim normal sapma olarak adlandırılır. Standardize artıklarda mutlak değerce 3'den büyük olan standartlaştırılmış hata terimi değerleri aykırı değer olarak kabul edilir.

RES_1	SRE_1
-2,94565	-,60896
4,03259	1,04778
-2,13484	-,44467
3,16896	,72880
-7,69494	-1,71042
-1,32041	-,29043
2,74774	,65101
6,93521	1,71111
-8,15290	-1,68084
-1,37118	-,33692
4,67362	1,05539
-2,12843	-,56765
-,06438	-,01323
4,25461	1,05683

Student Türü Artıklar
(Studentized Residual:SRE)

$$r_i = \frac{e_i}{\sqrt{V(e_i)}} = \frac{e_i}{\sqrt{AKO(1-h_{ii})}}$$

Student türü artıklar [-2,+2] aralığında yer alırlar. Mutlak değerce 2’den büyük student artık değer aykırı değer olarak kabul edilir.

$$h_{ii} = X_i{}'(X'X)^{-1}X_i$$

RES_1	DRE_1
-2,94565	-3,38605
4,03259	7,32249
-2,13484	-2,49120
3,16896	4,50815
-7,69494	-10,22593
-1,32041	-1,71825
2,74774	4,14866
6,93521	11,35528
-8,15290	-9,32060
-1,37118	-2,22673
4,67362	6,41024
-2,12843	-4,07191
-,06438	-,07311
4,25461	7,06076

Çıkartılmış Artıklar
(Deleted Residual-DRE)

$$e_{(i)} = y_i - \hat{y}_{i(i)}$$

i.gözlem çıkartıldıktan sonra geriye kalan n-1 gözlem üzerine regresyon modeli kurulup bu model yardımıyla i.nci gözlemin kestirim değerinin elde dilmesiyle bulunur.
Bu artık PRESS artığı olarak adlandırılır.

$$e_{(i)} = \frac{e_i}{1-h_{ii}}$$

RES_1	SDR_1
-2,94565	-,58873
4,03259	1,05352
-2,13484	-,42608
3,16896	,71052
-7,69494	-1,92920
-1,32041	-,27670
2,74774	,63112
6,93521	1,93029
-8,15290	-1,88254
-1,37118	-,32146
4,67362	1,06213
-2,12843	-,54741
-,06438	-,01255
4,25461	1,06376

Studentized Deleted Residual (SDR)

y-ekseni etrafındaki kuşkulu gözlemleri (aykırı değerleri) bulmada kullanılır (Hadi and Simonoff, 1993).

$$S^2_{(i)} = \frac{AKT_{(i)}}{n-k-2}$$

MAH_1
,76226
4,91216
,93104
2,93320
2,28903
2,08144
3,46126
4,13171
,70009
4,06629
2,59331
5,27619
,62403
4,23801

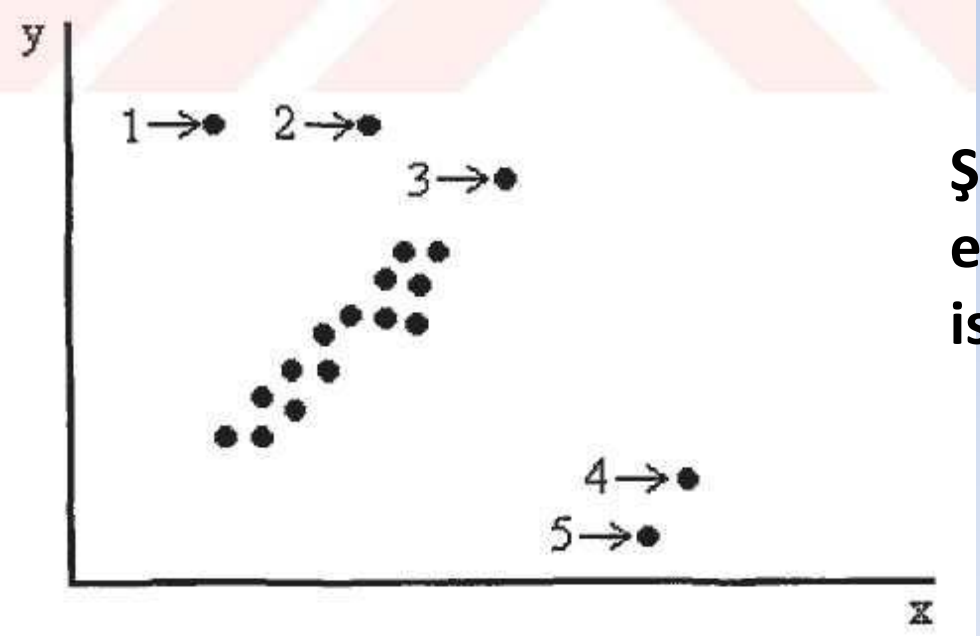
Mahalanobis Uzaklığı (MU)

X ekseni etrafındaki aykırı değerlerin bulunmasında kullanılır. Mu değerlerinin Box plot grafiği çizilir ve aykırı değer olup olmadığı belirlenir.

ETKİLİ GÖZLEMLER(influential observations)

Regresyon çözümlemesinden çıkartıldığında hesaplanan çeşitli değerleri (katsayılar, standart hatalar, t-F değerleri gibi) önemli ölçüde değiştiren gözlemlere etkili gözlemler denir.

Etkili gözlemler hem y-ekseni hem de x-ekseni yönünde kuşkulu gözlemler olabilmektedir.



Şekilde 1,4 ve 5 nolu gözlemler etkili gözlemlerdir. 2 nolu gözlem ise araştırılmalıdır.

ETKİLİ GÖZLEMLER(influentialobservations)

Bir gözlemin etkisini incelemek için "gözlemi çıkartma" ilkesi uygulanır. Regresyon modelini oluşturan veri kümesinden herbir gözlemin çıkartılması yoluyla elde edilen modellerdeki çeşitli değerlerin veri kümesinin oluşturduğu modeldeki değerlerle karşılaştırılması esasına dayanır. Hangi gözlemlerde değişim büyük olursa o gözlemler modeli büyük ölçüde etkiler (Aşıkgil, 2006).

Etkili gözlemler aşağıdaki testlerle de bulunabilir.

-**DFFITS İstatistiği :** $|DFFITS| > 2\sqrt{p/n}$ p=bağ.değ.say+1

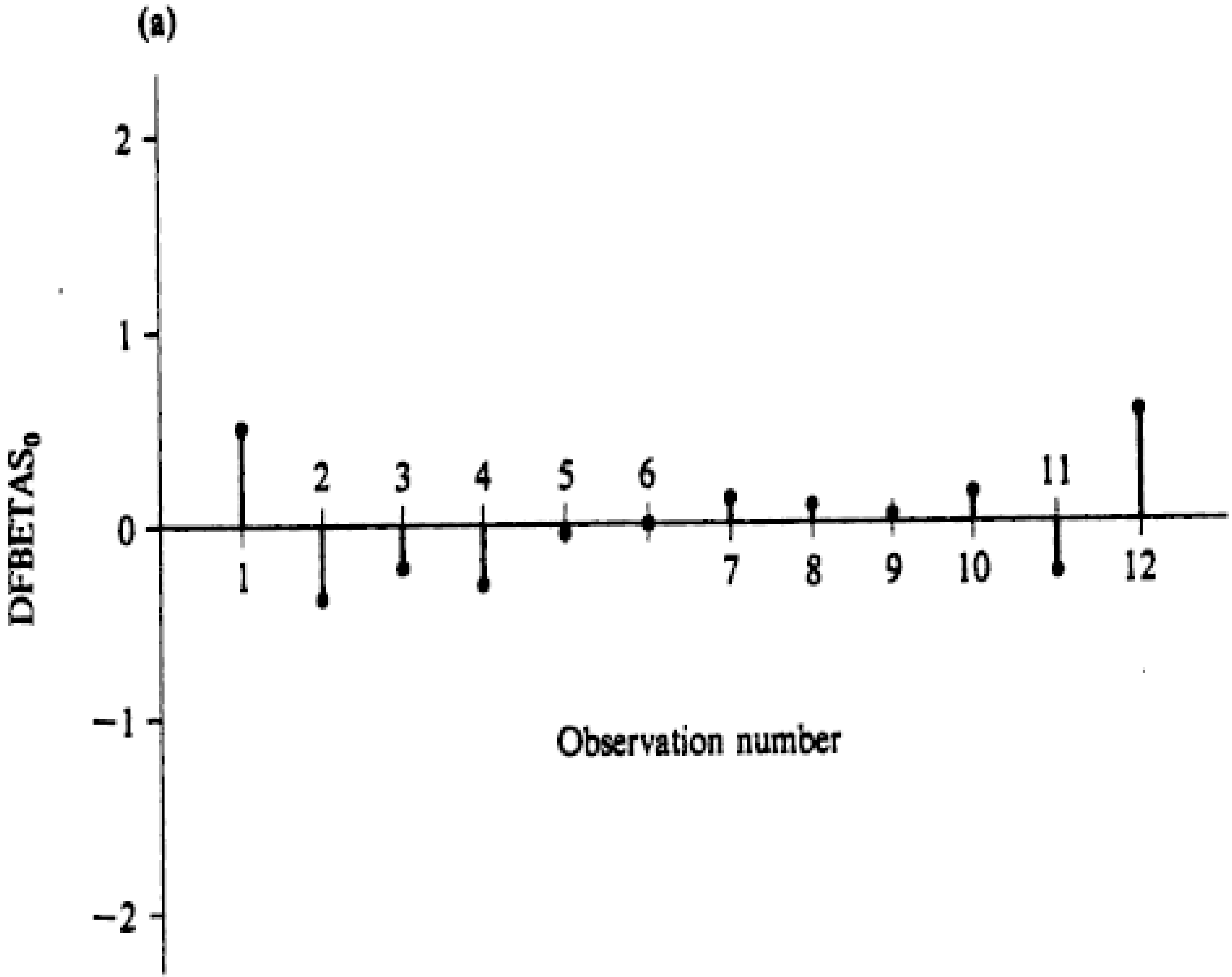
- **DFBETAS İstatistiği :** $|DFBETAS| > 2/\sqrt{n}$

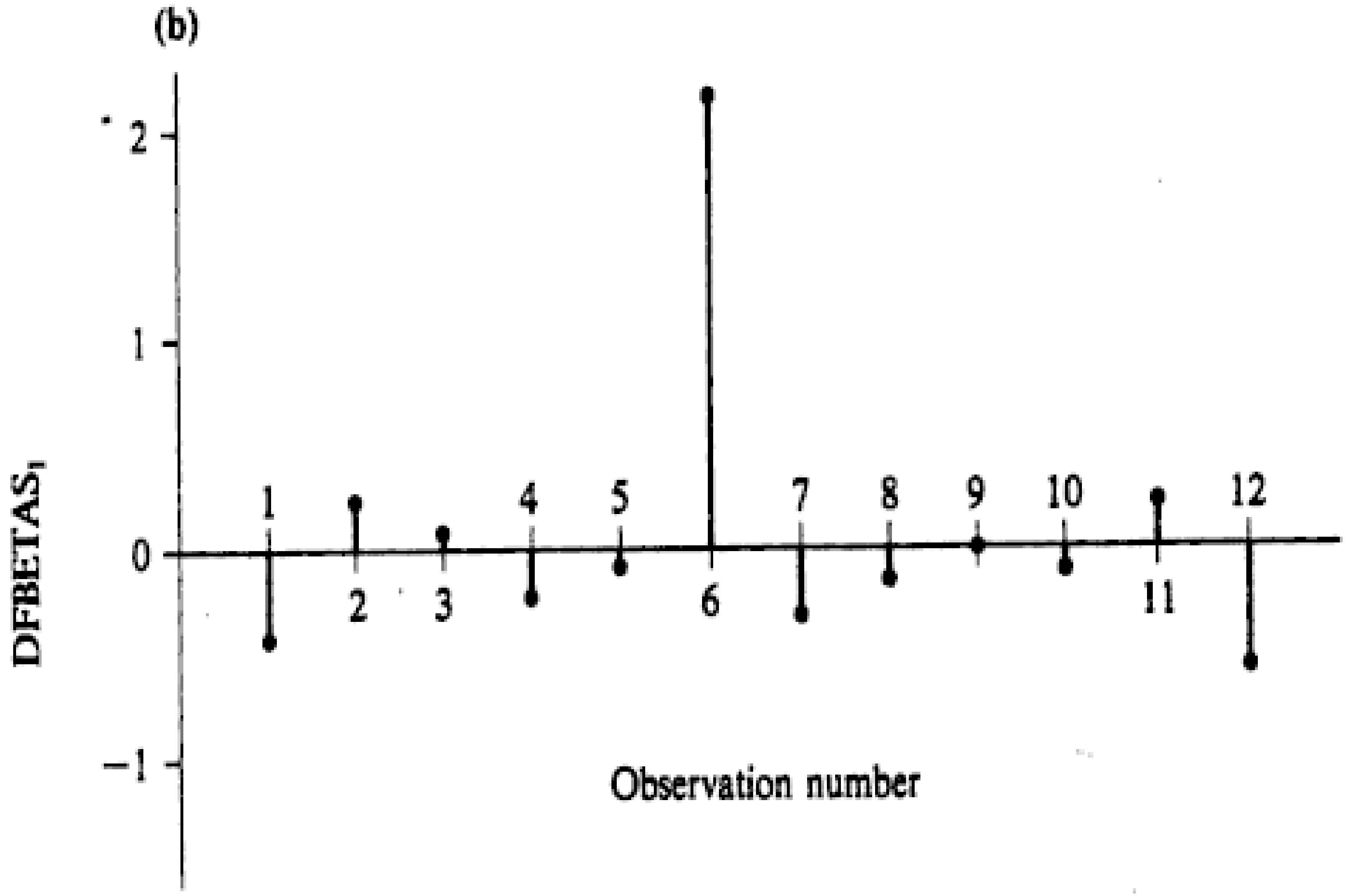
- **COVRATIO İstatistiği** :COVRATIO>1

Bu değerler gözlemlerin tahmini regresyon katsayıları ve tahmini y değerleri üzerindeki etkileri belirler. Kestirimin kesinliği ile ilgili bir bilgi vermezler.

Influence Measure	*Formula*	*Observation i May Be Influential If:*
Cook's D_i	$\dfrac{(\widehat{\boldsymbol{\beta}}_{(i)}-\widehat{\boldsymbol{\beta}})'(\boldsymbol{X}'\boldsymbol{X})(\widehat{\boldsymbol{\beta}}_{(i)}-\widehat{\boldsymbol{\beta}})}{p's^2}$	$D_i > F_{(.5,p',n-p')}$
DFFITS_i	$\dfrac{\widehat{Y}_i-\widehat{Y}_{i(i)}}{s_{(i)}\sqrt{v_{ii}}}$	$\lvert\text{DFFITS}_i\rvert > 2\sqrt{p'/n}$
Atkinson's C_i	$\left(\frac{n-p'}{p'}\right)^{1/2}\lvert\text{DFFITS}_i\rvert$	$\lvert C_i\rvert > 2[(n-p')/n]^{1/2}$
$\text{DFBETAS}_{j(i)}$	$\dfrac{\widehat{\beta}_j-\widehat{\beta}_{j(i)}}{s_i\sqrt{c_{jj}}}$	$\lvert\text{DFBETAS}_{j(i)}\rvert > 2/\sqrt{n}$
COVRATIO_i	$\dfrac{\det(s^2_{(i)}[\boldsymbol{X}'_{(i)}\boldsymbol{X}_{(i)}]^{-1})}{\det(s^2[\boldsymbol{X}'\boldsymbol{X}]^{-1})}$	$\text{COVRATIO}\begin{cases} < 1-3p'/n \\ > 1+3p'/n \end{cases}$

P’: Açıklayıcı değişken sayısı+1


(a)
2
1
0
−1
−2
DFBETAS$_0$
1
2
3
4
5
6
7
8
9
10
11
12
Observation number


(b)
DFBETAS$_i$
2
1
0
−1
1
2
3
4
5
6
7
8
9
10
11
12
Observation number

COO_1
,01386
,22391
,00825
,05611
,24056
,00635
,05402
,46651
,10116
,01771
,10347
,07356
,00001
,18416

Cooks Uzaklığı :

Etkili gözlemlerin (influential) elde edilmesinde kullanılır. Cook (1977) değeri 1 ve daha büyük olan gözlemler etkili gözlem olarak nitelenir.

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})'(X'X)(\hat{\beta} - \hat{\beta}_{(i)})}{df\,(AKO)}$$

LEV_1
,05864
,37786
,07162
,22563
,17608
,16011
,26625
,31782
,05385
,31279
,19949
,40586
,04800
,32600

Kaldıraç (Leverage):

Etkili gözlem elde edilmesinde kullanılan bir diğer yöntemdir.

Kaldıraç bağımlı değişkenin gözlenen değerinin tahmin edilen değer üzerindeki etkisini açıklar.

i, regresyondaki bağımsız değişken sayısı ise, merkezi kaldıracın ortalama değeri p/n olur. Sıfır olan kaldıraç noktanın regresyon doğrusunun uyumu ile ilgili olarak hiç etkisi olmadığını ifade eder. Kaldıraçların p/n'e yakın olması arzu edilir.

Kaldıracı 2p/n'den büyük olan noktalar incelenmelidir.

2*4/14=0,57'den büyük leverage değeri yoktur.

PRESS İSTATİSTİĞİ

PRESS istatistiği regresyon modellerini karşılaştırmada kullanılabilir. Küçük PRESS değerli model tercih edilir. PRESS bir regresyon modelininin yeni verilerin önkestirimlerinde ne kadar iyi performans göstereceğinin bir ölçütü olarak kabul edilir.

$$\mathrm{PRESS} = \sum_{i=1}^{n} (y_i - \hat{y}_{i,-i})^2$$

Örnek. Takım tezgahları üretimi yapan bir firma, satışlarını artırmak üzere neler yapılması gerektiğinin belirlenmesi amacıyla yaptığı bir araştırmada satış gelirlerini (Y) etkileyebilecek en önemli değişkenlerin bayi giderleri (X1), reklam giderleri (X2) ve rakip bayi sayısı (X3) olabileceğini belirlemiştir. Bu amaçla rasgele seçilen 10 adet bayiden elde edilen rakamlar aşağıdaki gibidir. Regresyon modelini tahmin ediniz ve açıklayıcı değişkenlerden hangisinin modele ek katkısı olduğunu bulunuz.

Y	X1	X2	X3
90	50	10	4
100	60	13	5
102	42	10	3
118	35	11	2
125	40	12	3
135	45	13	2
145	60	11	3
160	55	8	1
175	61	14	1
190	72	18	1

Coefficients[a]

	B	Std. Error	t	Sig.	95% Confidence Interval for B		Correlations		
					Lower Bound	Upper Bound	Zero-order	Partial	Part
(Constant)	98,99659	19,57757	5,057	,002	51,092	146,901			
X1	1,21202	,34120	3,552	,012	,377	2,047	,622	,823	,347
X2	1,45119	1,45315	,999	,357	-2,105	5,007	,529	,378	,097
X3	-18,17442	2,48839	-7,304	,000	-24,263	-12,086	-,842	-,948	-,713

a. Dependent Variable: Y

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9492,26301	3	3164,08767	32,97430	,000[a]
	Residual	575,73699	6	95,95616		
	Total	10068,00000	9			

Değişim Kaynağı	sd	Type I (Sequential) SS=KT	Type III (Partial) SS=KT	MS=KO	F
Regresyon	3	9492,1	9492,1	3164,03	F=32,97
X_1	1	3889,3	1210,6	1210,6	F_1=12,61
X_2	1	484,3	95,5	95,5	F_2=0,990
X_3	1	5118,6	5118,6	5118,6	F_3=53,33
Artık	6	575,9			
Genel	9	10068			

Değişim Kaynağı	sd	Type I (Sequential) SS=KT	Type III (Partial) SS=KT	MS=KO	F
Regresyon	3	9492,1	9492,1	3164,03	F=32,97
X_1	1	3889,3	1210,6	1210,6	F_1=12,61
X_2	1	484,3	95,5	95,5	F_2=0,990
X_3	1	5118,6	5118,6	5118,6	F_3=53,33
Artık	6	575,9			
Genel	9	10068			

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

1. F1=12,61>F1,6,0.05=5,99 olduğundan H0 red edilir. X1'in modele anlamlı bir katkısı vardır ve modelde yer almalıdır.

2. F2=0,99<F1,6,0.05=5,99 olduğundan H0 red edilemez. X2'in modele anlamlı bir katkısı yoktur ve modelde yer almamalıdır.

3. F3=53,33>F1,6,0.05=5,99 olduğundan H0 red edilir. X3'in modele anlamlı bir katkısı vardır ve modelde yer almalıdır.

Soru. Bir süt sığırcılığı çalışmasında laktasyon süt verimine (kg) etki ettiği düşünülen, yaş (ay), ırk (esmer= 0, jersey=1), günlük sağım sayısı, ve mevsim (kış ve sonbahar=0, ilkbahar ve yaz=1) özellikleri 33 hayvan üzerinde inceleniyor. Buna göre en iyi regresyon setini bulunuz.

No	LSV	YAŞ	M	SS	IRK	No	LSV	YAŞ	M	SS	IRK
1	2850	41	0	2	0	**18**	4500	93	1	3	0
2	3415	65	0	4	0	**19**	3400	105	0	3	0
3	3654	77	1	2	0	**20**	3560	42	1	2	0
4	3250	101	1	2	0	**21**	3330	53	0	2	1
5	3500	42	1	3	0	**22**	3450	66	0	3	1
6	3854	69	1	2	0	**23**	3900	77	1	2	0
7	3752	81	1	3	0	**24**	4352	91	1	4	0
8	3400	89	0	2	0	**25**	3115	44	0	2	1
9	3978	101	1	3	0	**26**	3200	53	0	2	1
10	2954	41	0	2	0	**27**	3900	65	1	3	0
11	3600	66	1	3	0	**28**	4212	79	1	3	0
12	3897	80	1	4	0	**29**	3600	91	1	4	0
13	4050	90	0	3	0	**30**	3550	45	1	2	0
14	3762	103	0	4	0	**31**	3265	57	0	2	1
15	3100	43	1	3	0	**32**	3115	68	0	3	1
16	3356	53	1	3	0	**33**	4522	77	1	2	0
17	3800	78	1	4	0						