

BÖLÜM VII LINEER REGRESYON ANALİZİ

VII.1 Linear Regresyon Analizi

Bir bağımlı değişken ile bir ya da daha fazla bağımsız değişken arasındaki ilişki bir matematiksel model ile ifade edilmek istenebilir. Eğer söz konusu matematiksel model bir doğrusal fonksiyon ise bu **modele Linear Regresyon Modeli** adı verilir. Bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin doğrusal olup olmadığı Serpme diyagramı ile görsel olarak belirlenebilir. Bir bağımlı ve bir bağımsız değişkenin yer aldığı lineer regresyon modeline **Basit Linear Regresyon Modeli**, bir bağımlı ve birden fazla bağımsız değişkenin bulunduğu lineer regresyon modeline ise **Çoklu Linear Regresyon Modeli** denir.

Bağımlı değişken; değeri başka değişkenler tarafından belirlenen ve diğer değişkenlerin değeri değiştiğinde bu değişimden etkilenen değişkene denir. Bu sebeple etkilenen, açıklanan veya cevap değişkeni olarak da adlandırılır. Genel olarak Y ile gösterilir.

Bağımsız değişken(ler); Değeri rastgele koşullara göre oluşan, bağımsız olarak değişim gösteren ve başka değişkenlerin değişimi üzerine etkide bulunan değişken(ler)e denir. Bu sebeple etkileyen, açıklayan değişken veya faktör olarak da ifade edilir. Bağımsız değişken tek olduğu zaman X sembolü ile, birden fazla sayıda (örneğin k tane) ise X_1, X_2, \dots, X_k sembolleri ile gösterilir.

Basit Linear Regresyon Modeli: Bir bağımlı değişken (Y) ile bir bağımsız değişken (X) arasındaki neden-sonuç ilişkisini

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (7.1)$$

şeklinde bir lineer fonksiyon olarak ifade eden modele denir. Burada β_0 ve β_1 modelin bilinmeyen parametreleri ve ε hata terimidir.

Çoklu Linear Regresyon Modeli: Bir bağımlı değişken (Y) ile birden fazla (örneğin k tane) bağımsız değişken (X_1, X_2, \dots, X_k) arasındaki neden-sonuç ilişkisini

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (7.2)$$

şeklinde bir lineer fonksiyon olarak ifade eden modele denir. Burada β_0, β_1, \dots ve β_k modelin bilinmeyen parametreleri ve ε hata terimidir.

Varsayımları:

1. Bağımlı ve bağımsız değişkenler nicel, kesikli/sürekli türden olmalı ve eşit aralıklı veya oranlama düzeyinde ölçülmelidir.
2. Bağımlı değişken normal dağılım göstermelidir [$Y \sim N(\mu_Y, \sigma_Y^2)$]

3. Bağımsız değişkenler hatasız ölçülmeli ve aralarında çoklu doğrusal bağımlılık (çoklu bağlantı) olmamalıdır..

4. Hata terimleri birbirinden bağımsız, sıfır ortalamalı ve σ^2 varyanslı (aynı varyanslı) normal dağılım göstermelidir [$\varepsilon \sim N(0, \sigma^2)$]

5. Veri grubunda gözlemler arasında ardışık bağımlılık (otokorelasyon) bulunmamalıdır.

Amaç:

1. Eşitlik (1) veya Eşitlik (2) ile verilen modellerin parametrelerini tahmin etmek. Bunun için En Küçük Kareler (EKK) yöntemi kullanılır. $j=0,1,2,\dots,k$ için β_j parametresinin tahmin edicisi b_j istatistiği ile gösterilir.

2. Model denklemini (regresyon denklemini) tahmin etmek.

Basit lineer regresyon modelinin tahmin denklemi $\hat{Y} = b_0 + b_1X$ iken, çoklu lineer regresyon modelinin tahmin denklemi $\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$ şeklindedir. Tahmin denklemi hata terimi kapsamaz.

3. Model parametreleri hakkında istatistiksel çıkarımlar yapmak. Bunun için model parametreleri ile ilgili güven aralıkları oluşturulur ve parametrelerle ilgili iddialar hipotez testi ile test edilir. Bu amaçla test edilecek hipotezler:

$$H_0: \beta_j = 0$$

$H_1: \beta_j \neq 0$, ($j=1,2,\dots,k$) şeklinde kurulur. Bu hipotezlerin test edilmesinde test istatistiği olarak t_{n-k-1} şeklinde gösterilen $(n-k-1)$ serbestlik dereceli student t-istatistiği kullanılır. Karar kuralı $p < \alpha$ ise H_0 ret edilir, $p \geq \alpha$ ise H_0 ret edilemez.

4. Model denkleminin anlamlılığını (doğrusallığa uyumun önemliliğini) kontrol etmek. Bunun için test edilecek hipotezler;

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \text{ (Doğrusallığa uyum önemsiz)}$$

$$H_1: \text{En az bir } \beta_j \neq 0 \text{ (Doğrusallığa uyum önemli)}$$

şeklinde kurulur. Test işlemini sonuçları aşağıdaki gibi düzenlene varyans analizi tablosunda özetlenebilir.

Tablo 6.14 Model Uyumluluğu ANOVA Tablosu

Kaynak	s.d.	KT	KO	Test istatistiği
Regresyon (R)	k	RKT	$RKO = \frac{RKT}{k}$	$F = \frac{RKO}{AKO} \sim F_{k;n-k}$
Artık (A)	n-k	AKT	$AKO = \frac{AKT}{n-k}$	$p < \alpha$ ise H_0 ret edilir
Genel	n-1	KT_Y	-----	$p \geq \alpha$ ise H_0 kabul edilir.

H_0 hipotezi kabul edilirse doğrusallığa uyum önemsiz, ret edilirse doğrusallığa uyum önemlidir.

5. Modelin geçerliliği değerlendirilir. Bunun için model geçerliliği ile ilgili bazı kriterlere bakmak gerekir. Bu kriterler:

a) Belirtme / ÇokluBelirtme katsayısı (R^2): Modelin belirleyicilik gücünü ifade eder. Bağımsız değişkenlerin bağımlı değişkeni açıklama oranıdır. Modele bağımsız değişken ekledikçe değeri artar.

$$R^2 = \frac{RKT}{KT_Y} \quad (7.3)$$

eşitliği ile hesaplanır. Bu oranın yüksek olması istedir. (%80 den fazla)

b) Denklemin Standart hatası ($\hat{\sigma}$): Modelin geçerli bir model olması için denklemin standart hatasının küçük olması gerekir. Denklemin standart hatası, belirtme katsayısı ile ters orantılıdır. Belirtme katsayısı artarken standart hata azalır. Denklemin standart hatası:

$$\hat{\sigma} = \sqrt{AKO} \quad (7.4)$$

eşitliği ile bulunur.

c) Varyans şişkinlik faktörü (VIF): Bağımsız değişkenler arasında çoklu bağlantı olup olmadığını test etmek için kullanılan bir istatistiktir. VIF büyük değer alıyorsa regresyon katsayılarının (b_j 'lerin) varyansları büyür ve bağımsız değişkenlerin bağımlı değişken üzerine etkileri yanlış değerlendirilir. Bu durumda regresyon modeli geçersiz olur. VIF değerinin değerlendirilmesi:

- Eğer tüm bağımsız değişkenler için $VIF=1$ ise çoklu bağlantı yoktur
- $1 < VIF \leq 5$ ise orta düzeyde çoklu bağlantı vardır. Modelde düzeltme yapmak gerekmez
- $5 < VIF \leq 10$ ise yüksek düzeyde çoklu bağlantı vardır. Bağımsız değişkenlerde düzeltme yapılmalıdır. Önemsiz ilişki düzeylerine sahip değişkenler sırası ile teker teker modelden çıkartılır ve analiz yenilenir. Çoklu bağlantıya neden olan değişken bulunarak düzeltilir ya da modelden çıkartılır ve analiz yenilenir.
- $VIF > 10$ ise çok yüksek düzeyde önemli çoklu bağlantı vardır. Bu durumda model geçersizdir. Uygun yöntemlerle bağımsız değişkenler düzeltilir yeniden model oluşturulur.

d) Durbin-Watson (DW) Testi: Ardışık gözlemler ya da hata terimleri arasında otokorelasyon olup olmadığını belirler. Eğer otokorelasyon varsa, regresyon katsayılarının varyansları çok küçük çıkar ve önemli etkiye sahip olmayan bağımsız değişkenler önemli etkiye sahip gibi değerlendirilir. DW istatistiği $0 \leq DW \leq 4$ aralığında değer alır. Durbin-Watson testi ile otokorelasyonun olup olmadığı genel anlamda test edilebileceği gibi, pozitif otokorelasyonun ya da negatif otokorelasyonun olup olmadığı da ayrı ayrı kontrol edilebilir.

Eğer genel anlamda otokorelasyon olup olmadığı araştırılacaksa test edilecek hipotezler;

$$H_0: \rho = 0 \text{ (Otokorelasyon yok)}$$

$$H_1: \rho \neq 0 \text{ (Otokorelasyon var)} \quad (7.5)$$

şeklinde. SPSS çözümü sonucunda $DW=0$ ise pozitif otokorelasyon var, $DW=2$ ise otokorelasyon yok ve $DW=4$ ise negatif otokorelasyon vardır. Ancak DW istatistiğinin, D_L alt sınır ve D_U üst sınır olmak üzere bu iki sınır değeri arasında yer almaktadır. Eğer DW değeri bu sınırlar dışında ise yukarıdaki hipotezler için bir sonuca varılabilir. Burada D_L ve D_U alt ve üst kritik değerler olup, n örnek hacmi, k bağımsız değişken sayısı ve α önem seviyesi olmak üzere bunlara bağlı olarak Durbin-Watson test istatistiği tablosundan bulunabilir. Bu durumda pozitif ya da negatif otokorelasyon varlığı ayrı ayrı incelenebilir. Pozitif otokorelasyon için test edilecek hipotezler:

$$H_0: \rho = 0 \text{ (pozitif otokorelasyon yok)}$$

$$H_1: \rho > 0 \text{ (pozitif otokorelasyon var)} \quad (7.6)$$

şeklinde iken, negatif otokorelasyon için test edilecek hipotezler:

$$H_0: \rho = 0 \text{ (negatif otokorelasyon yok)}$$

$$H_1: \rho < 0 \text{ (negatif otokorelasyon var)} \quad (7.7)$$

şeklinde oluşturulur. Karar kuralı ise aşağıdaki gibidir:

$0 < DW < D_L$ ise (7.6) daki H_0 hipotezi ret edilir ve pozitif otokorelasyon vardır.

$D_U < DW < 4 - D_U$ ise (7.6) daki H_0 hipotezi ret edilemez ve otokorelasyon yoktur.

$D_L \leq DW \leq D_U$ ise (7.6) daki H_0 hipotezi ile ilgili sonuç yetersiz (otokorelasyon hakkında karar verilemez)

$4 - D_L < DW < 4$ ise (7.7) daki H_0 hipotezi ret edilir ve negatif otokorelasyon vardır.

$D_U < DW < 4 - D_U$ ise (7.7) daki H_0 hipotezi ret edilemez ve otokorelasyon yoktur.

$4 - D_U \leq DW \leq 4 - D_L$ ise (7.7) daki H_0 hipotezi ile ilgili sonuç yetersiz (otokorelasyon hakkında karar verilemez).

SPSS ile Regresyon Analizi Uygulama Algoritması

Adım.1 Tüm değişkenler tanımlanır ve özellikleri girilir

Adım.2 Veriler değişkenlere göre sütunlara girilir

Adım.3 Bağımlı değişken ile bağımsız değişken(ler)in serpmeye diyagramı incelenerek ilişkinin doğrusal olup olmadığı belirlenir.

Adım.4 **Analyze > Regression > Linear** seçenekleri seçilir. Açılan ekranda değişkenler listesinden bağımlı değişken seçilerek **Dependent** işlem kutusuna, bağımsız değişken(ler) seçilerek **Independent(s)** işlem kutusuna aktarılır.

Adım.5 **Statistics** penceresi açılarak, açılan ekrandan **Estimates, Confidence intervals, Model fit, Descriptives, Collinearity diagnostics** ve **Durbin-Watson** seçenekleri seçilir. **Continue** tıklanarak pencere kapatılır.

Adım.6 **Ok** tıklanır ve işlem bitirilir. Sonuçlar **Output** sayfasında tablolar halinde görüntülenir. Bu tablolar değerlendirilir ve sonuçlar yorumlanır.

ÖRNEK 7.1 Rastgele seçilen 16 bireyin yaş, boy, ağırlık, GİSS(günlük içilen sigara sayısı) ve SKB(sistolik kan basıncı) değerleri aşağıdadır. Buna göre :

- Değişkenlerin özelliklerini belirtiniz? Hangi değişken, Hangi değişken/değişkenler tarafından açıklanabilir?
- SKB ve Yaş değişkenleri için SPSS programını kullanarak bu iki değişken arasındaki doğrusal ilişkinin derecesini ve önemli olup olmadığını %5 önem seviyesinde kontrol ediniz?
- SKB ve Yaş değişkenleri arasındaki doğrusal ilişkiye ait basit lineer regresyon modelini ifade ediniz, SPSS programı yardımı ile modelin tahmin denklemini elde ediniz ve sonuçları değerlendiriniz?
- SKB ve GİSS değişkenleri için SPSS programını kullanarak bu iki değişken arasındaki doğrusal ilişkinin derecesini ve önemli olup olmadığını %5 önem seviyesinde kontrol ediniz?
- SKB ve GİSS değişkenleri arasındaki doğrusal ilişkiye ait basit lineer regresyon modelini ifade ediniz, SPSS programı yardımı ile modelin tahmin denklemini elde ediniz ve sonuçları değerlendiriniz?
- Y: SKB ve X_1 : Yaş, X_2 : Boy, X_3 : ağırlık, X_4 : GİSS olmak üzere bu değişkenler arasındaki çoklu lineer regresyon modelini ifade ediniz, SPSS programı yardımı ile modelin tahmin denklemini elde ediniz ve sonuçları değerlendiriniz?

Tablo 7.1 Bireylerin Bazı Özelliklerine Ait Ölçümler

Birey	SKB	Yaş	Boy	Ağırlık	GİSS	Birey	SKB	Yaş	Boy	Ağırlık	GİSS
1	115	51	166	67	10	9	110	48	165	70	0
2	122	64	165	61	15	10	119	78	152	58	10
3	130	46	174	83	20	11	130	39	177	82,5	18
4	126	39	168	78,9	25	12	120	51	166	63	20
5	110	58	162	67	0	13	149	73	178	93,1	45
6	141	54	178	90	30	14	125	53	174	89	27
7	124	31	171	77,7	12	15	125	56	169	72	30
8	150	67	173	89,3	40	16	114	47	159	64	5

Cözüm a) Yaş: Nicel türden sürekli bir değişken olup ölçme düzeyi oranlama

Boy: Nicel türden sürekli bir değişken olup ölçme düzeyi oranlama

Ağırlık: Nicel türden sürekli bir değişken olup ölçme düzeyi oranlama

GİSS: Nicel türden kesikli bir değişken olup ölçme düzeyi eşit aralıklı

SKB: Nicel türden sürekli bir değişken olup ölçme düzeyi oranlama

Ayrıca SKB, diğer değişkenlerden etkilenen bir değişken olduğundan bağımlı/açıklanan değişken olarak alınır. Bunu etkileyen diğer değişkenler ise bağımsız/açıklayan değişkenler olarak alınır.

b) Nicel türden sürekli ve ölçme düzeyi eşit aralıklı veya oranlama olan değişkenler arasındaki doğrusal ilişkinin hangi korelasyon katsayısı ile ölçüleceğine karar verebilmek için değişkenlere ait verilerin normal dağılım ile uyumlu olup olmadığının kontrol edilmesi gerekir. Bu amaçla test edilecek hipotezler:

H_0 : SKB ve Yaş değişkenlerine ait dağılımlar, normal dağılım ile uyumlu

H_1 : SKB ve Yaş değişkenlerine ait dağılımlar, normal dağılım ile uyumlu değil

SPSS ile yapılan test işlemi sonuçları Tablo 7.2’de verildi.

Tablo:7.2 Normallik Testi Sonuçları

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SKB	,175	16	,200*	,909	16	,113
Yaş	,108	16	,200*	,977	16	,940
Boy	,123	16	,200*	,949	16	,480
Ağırlık	,140	16	,200*	,934	16	,279
GİSS	,100	16	,200*	,967	16	,786

Tablo7.2’ye göre hem SKB hem de Yaş değişkenlerine ait verilerin dağılımı normal dağılım ile uyumludur. Çünkü önem seviyesi $\alpha = 0,05$ olmak üzere SKB için $p = 0,113 > \alpha$ ve Yaş için $p = 0,940 > \alpha$ olup H_0 hipotezi ret edilemez. Bu sebeple bu iki değişken arasındaki doğrusal ilişkinin değerlendirilmesi Pearson korelasyon katsayısı ile yapılmalıdır. Doğrusal ilişkinin önemliliği için test edilecek hipotezler:

$H_0: r = 0$ (doğrusal ilişki önemsiz)

$H_1: r \neq 0$ (doğrusal ilişki önemli)

şeklinde dir. SPSS test işlemi sonucu Tablo7.3 ile verildi.

Tablo 7.3 SKB ve Yaş Arasındaki İlişki İncelemesi

		SKB	Yaş
SKB	Pearson Correlation	1	,285
	Sig. (2-tailed)		,284
	N	16	16
Yaş	Pearson Correlation	,285	1
	Sig. (2-tailed)	,284	
	N	16	16

Tablo 7.3'e göre SKB ile Yaş değişkenleri arasındaki korelasyon $r = 0,285$ olup, değişkenler arasında aynı yönde ama zayıf bir doğrusal ilişki vardır. $p = 0,284$ olup $\alpha = 0,05$ önem seviyesinde $p > \alpha$ olduğundan H_0 hipotezi ret edilemez. Buna göre SKB ve Boy değişkenleri arasındaki doğrusal ilişki önemsizdir.

c) SKB ve Yaş değişkenleri arasındaki doğrusal ilişkinin modeli için bağımlı değişken $Y: SKB$ ve bağımsız değişken $X = Yaş$ olmak üzere basit lineer regresyon modeli:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

ve tahmin denklemi :

$$\hat{Y} = b_0 + b_1 X$$

olup, SPSS çözümü Tablo 7.4 ile verildi.

Tablo 7.4 Regresyon Denklemi ve Regresyon Katsayılarının Değerlendirilmesi

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	110,784	13,665	8,107	,000	81,475	140,094		
	Yas	,278	,249	,285	1,114	-,257	,813	1,000	1,000

Tablo 7.4'e göre modelinin tahmin denklemi, $\hat{Y} = b_0 + b_1 X = 110,784 + 0,278X$ şeklinde elde edilir. Bu denkleme göre yaştaki 1 birimlik değişim (artma veya azalma) SKB değerinde %27,8 oranında bir değişime (artma ya da azalmaya) yol açacaktır.

Model parametrelerinin değerlendirilmesi: Model parametrelerinden β_1 doğrunun eğim parametresi iken, β_0 ise doğrunun y-eksenini kestiği noktaya gösteren parametredir. Bu parametreler bilinmez, ancak örneklemden tahmin edilirler. β_0 parametresinin tahmin edicisi

$b_0 = 110,784$ olup tahmin doğrusu y-eksenini bu noktada kesmekte ve $x=0$ iken Y 'nin tahmin değerinin $\hat{Y} = b_0 = 110,784$ olduğunu gösterir. Bu ise sıfır yaşındaki bir bireyin SKB değerinin 110,784 birim olduğu anlamına gelir (**Not:** Genellikle β_0 ve b_0 üzerinde fazla durulmaz, çünkü bazen anlamsız bir değerlendirme verebilir). Burada esas olan β_1 parametresi ve bunun tahmin edicisi olan b_1 istatistiğidir. Bu parametrenin anlamlı olup olmadığı:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

hipotezlerinin test edilmesi ile belirlenir. Test işleminin sonucu Tablo 7.4'te yer almaktadır. Test istatistiğinin alabileceği değer $t_{n-k-1} = t_{16-1-1} = t_{14} = 1,114$ ($n = 16, k = 1$) ve $p = 0,284$ dır. Bu durumda $\alpha = 0,05$ için $p > \alpha$ olup H_0 ret edilemez. Bu sonuca göre β_1 parametresi önemsiz olup Yaş ile SKB arasında doğrusal bir ilişki anlamlı değildir. Ayrıca β_1 parametresi için güven aralığı $P(-0,257 \leq \beta_1 \leq 0,813)=0,95$ olup sıfırı kapsamaktadır. Bu durumda β_1 parametresi önemsiz olduğunun bir göstergesidir.

Model anlamlılığının testi: Basit lineer regresyon analizinde model anlamlılığının testi ile β_1 parametresinin değerlendirilmesi eşanlamlıdır. Burada test edilecek hipotezler:

$$H_0: \text{Modelin doğrusallığı uyumu anlamsız } (\beta_1 = 0)$$

$$H_1: \text{Modelin doğrusallığı uyumu anlamlı } (\beta_1 \neq 0)$$

Test işleminin sonuçları ANOVA tablosu olarak bilinen Tablo 7.5'de özetlenmiştir.

Tablo 7.5 Model Anlamlılığı İçin ANOVA Tablosu

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	182,636	1	182,636	1,241	,284 ^b
Residual	2061,114	14	147,222		
Total	2243,750	15			

Tablo 7.5'e göre test istatistiğinin alabileceği değer $F_{k;n-k-1} = F_{1;14} = 1,241$ ve $p = 0,284$ dır. $\alpha = 0,05$ için $p > \alpha$ olduğundan H_0 ret edilemez. Yani modelin doğrusallığı uyumu anlamsızdır. Bu sonuca göre yaş ile SKB arasındaki ilişki doğrusal bir modelle açıklanamaz.

Model geçerliliğinin değerlendirilmesi Modelin değerlendirilmesinde Tablo 7.6 ile çoklu bağlantı için Tablo 7.4'ün son kısmı dikkate alınır. Tablo 7.6'ya göre:

Tablo 7.6 Model Değerlendirmesi

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,285 ^a	,081	,016	12,13352	2,346

Belirtme katsayısı: $R^2 = 0,081$ olup çok düşüktür. Bu sonuca göre bağımlı değişkendeki değişimin (yani sistolik kan basıncına ait varyansın) %8,1 ini bağımsız değişken (yani Yaş) açıklamaktadır. Diğer bir ifade ile SKB'na ait varyansın $1-0,081=0,919$ yani %91,9'u bu model tarafından açıklanamamaktadır. Bu anlamda elde edilen doğrusal model iyi bir model değildir.

Denklemin standart hatası: Denklemin hata varyansı $AKO = 147,222$ olup, denklemin standart hatası $\hat{\sigma} = \sqrt{AKO} = \sqrt{147,222} = 12,133$ oldukça yüksektir. Bunun mümkün olduğu kadar küçük çıkması beklenir (en azından H_0 hipotezinin ret edilebileceği kadar). Denklemin standart hatasının büyük olması da doğrusal modelin iyi bir model olmadığının bir göstergesidir.

Varyans şişkinlik faktörü : $VIF=1$ (Tablo 7.4'den) Bağımsız değişken tek olduğu zaman bu değer hep “ 1 “ çıkacağından bu durumda genelde bakılmaz ve değerlendirmeye alınmaz. Çünkü çoklu bağlantı bağımsız değişkenler arasındaki ilişki ile ilgilidir. Bu sebeple modelde birden fazla bağımsız değişken olduğunda model değerlendirmesinde buna da yer verilir.

Durbin –Watson istatistiği: Tablo 7.6'ya göre $DW=2,346$ dır. $n = 16$, $k = 1$ ve $\alpha = 0,05$ iken DW tablosundan $D_L = 1,106$ ve $D_U = 1,371$ bulunmuştur. Buna göre $4-D_L = 4 - 1,106 = 2,894$ ve $4 - D_U = 4 - 1,371 = 2,629$ olup, $1,371 < 2,346 < 2,629$, yani $D_U < DW < 4 - D_U$ olduğundan $H_0: \rho = 0$ hipotezi kabul edilir ve otokorelasyon yoktur. Buna göre ardışık hata terimleri arasındaki ilişki önemsizdir.

d) Önce SKB ve GİSS değişkenlerine ait verilerin normal dağılım ile uyumlu olup olmadığının kontrol edilmesi gerekir. Bu amaçla test edilecek hipotezler:

H_0 : SKB ve GİSS değişkenlerine ait dağılımlar, normal dağılım ile uyumlu

H_1 : SKB ve GİSS değişkenlerine ait dağılımlar, normal dağılım ile uyumlu değil

Tablo 7.2'ye göre hem SKB hem de GİSS değişkenlerine ait verilerin dağılımı normal dağılım ile uyumludur. Çünkü önem seviyesi $\alpha = 0,05$ olmak üzere SKB için $p = 0,113 > \alpha$ ve GİSS için $p = 0,786 > \alpha$ olup H_0 hipotezi ret edilemez. Bu sebeple bu iki değişken arasındaki doğrusal ilişkinin değerlendirilmesi Pearson korelasyon katsayısı ile yapılmalıdır. Doğrusal ilişkinin önemliliği için test edilecek hipotezler:

$H_0: r = 0$ (doğrusal ilişki önemsiz)

$H_1: r \neq 0$ (doğrusal ilişki önemli)

şeklindeir. SPSS test işlemi sonucu Tablo 7.7 ile verildi.

Tablo 7.7 SKB ve GİSS Arasındaki İlişki İncelemesi

	SKB	GİSS
Pearson Correlation	1	,914**
SKB Sig. (2-tailed)		,000
N	16	16
Pearson Correlation	,914**	1
GİSS Sig. (2-tailed)	,000	
N	16	16

Tablo7.7’ye göre SKB ile GİSS değişkenleri arasındaki korelasyon $r = 0,914$ olup, değişkenler arasında aynı yönde ama güçlü bir doğrusal ilişki vardır. $p = 0,000$ olup $\alpha = 0,05$ önem seviyesinde $p > \alpha$ olduğundan H_0 hipotezi ret edilir. Buna göre SKB ve GİSS değişkenleri arasındaki doğrusal ilişki önemlidir.

e) SKB ve GİSS değişkenleri arasındaki doğrusal ilişkinin modeli için bağımlı değişken Y : SKB ve bağımsız değişken $X = GİSS$ olmak üzere basit lineer regresyon modeli:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

ve tahmin denklemi :

$$\hat{Y} = b_0 + b_1 X$$

olup, SPSS çözümü Tablo7.8 ile verildi.

Tablo 7.8 Regresyon Denklemi ve Regresyon Katsayılarının Değerlendirilmesi

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	109,348	2,314		47,251	,000		
GİSS	,848	,100	,914	8,448	,000	1,000	1,000

a. Dependent Variable: SKB

Tablo 7.8’e göre modelinin tahmin denklemi, $\hat{Y} = b_0 + b_1 X = 109,348 + 0,848X$ şeklinde elde edilir. Bu denkleme göre GİSS deki 1 birimlik değişim (artma veya azalma) SKB değerinde %84,8 oranında bir değişime (artma ya da azalmaya) yol açacaktır.

Model parametrelerinin değerlendirilmesi: β_0 parametresinin tahmin edicisi $b_0 = 109,348$ olup tahmin doğrusu y-eksenini bu noktada kesmekte ve $x=0$ iken Y ’nin tahmin değerinin $\hat{Y} = b_0 = 109,348$ olduğunu gösterir. Bu ise GİSS değeri sıfır olan bir bireyin SKB değerinin 109,348 birim olduğu anlamına gelir (**Not:** Genellikle β_0 ve b_0 üzerinde fazla durulmaz, çünkü bazen anlamsız bir değerlendirme verebilir). Burada esas olan β_1 parametresi ve bunun tahmin edicisi olan b_1 istatistiğidir. Bu parametrenin anlamlı olup olmadığı:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

hipotezlerinin test edilmesi ile belirlenir. Test işleminin sonucu Tablo 7.8’de yer almaktadır. Test istatistiğinin alabileceği değer $t_{n-k-1} = t_{16-1-1} = t_{14} = 8,488$ ($n = 16, k = 1$) ve $p = 0,000$ dır. Bu durumda $\alpha = 0,05$ için $p < \alpha$ olup H_0 ret edilir. Bu sonuca göre β_1 parametresi önemli olup GİSS ile SKB arasındaki doğrusal ilişki anlamlıdır.

Model anlamlılığının testi: Basit lineer regresyon analizinde model anlamlılığının testi ile β_1 parametresinin değerlendirilmesi eşanlamlıdır. Burada test edilecek hipotezler:

$$H_0: \text{Modelin doğrusallığı uyumu anlamsız } (\beta_1 = 0)$$

$$H_1: \text{Modelin doğrusallığı uyumu anlamlı } (\beta_1 \neq 0)$$

Test işleminin sonuçları ANOVA tablosu olarak bilinen Tablo 7.9’da özetlenmiştir.

Tablo 7.9 Model Anlamlılığı İçin ANOVA Tablosu

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	1875,769	1	1875,769	71,364	,000 ^b
Residual	367,981	14	26,284		
Total	2243,750	15			

a. Dependent Variable: SKB

b. Predictors: (Constant), GİSS

Tablo 7.5’e göre test istatistiğinin alabileceği değer $F_{k;n-k-1} = F_{1;14} = 71,364$ ve $p = 0,000$ dır. $\alpha = 0,05$ için $p < \alpha$ olduğundan H_0 ret edilir. Yani modelin doğrusallığı uyumu anlamlıdır. Bu sonuca göre GİSS ile SKB arasındaki ilişki doğrusal bir modelle açıklanabilir.

Model geçerliliğinin değerlendirilmesi Modelin değerlendirilmesinde Tablo 7.10 ile çoklu bağlantı için Tablo 7.8’in son kısmı dikkate alınır. Tablo 7.10’a göre:

Tablo 7.10 Model Değerlendirmesi

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,914 ^a	0,836	0,824	5,12683	1,617

Belirtme katsayısı: $R^2 = 0,836$ dir. Bu sonuca göre bağımlı değişkendeki değişimin (yani sistolik kan basıncına ait varyansın) %83,61 ini bağımsız değişken (yani GİSS) açıklamaktadır. Diğer bir ifade ile SKB’na ait varyansın $1-0,836=0,164$ yani %16,4’ü bu model tarafından açıklanamamaktadır. Bu anlamda elde edilen doğrusal model iyi bir modeldir.

Denklemin standart hatası: Denklemin hata varyansı $AKO = 26,284$ olup, denklemin standart hatası $\hat{\sigma} = \sqrt{AKO} = \sqrt{26,284} = 5,127$ olup düşüktür. Denklemin standart hatasının küçük olması da doğrusal modelin iyi bir model olduğunun bir göstergesidir.

Varyans şişkinlik faktörü : $VIF=1$ (Tablo 7.8'den) Bağımsız değişken tek olduğu zaman bu değer hep “ 1 “ çıkacağından bu durumda genelde bakılmaz ve değerlendirmeye alınmaz. Çünkü çoklu bağlantı bağımsız değişkenler arasındaki ilişki ile ilgilidir. Bu sebeple modelde birden fazla bağımsız değişken olduğunda model değerlendirmesinde buna da yer verilir.

Durbin –Watson istatistiği: Tablo 7.10'a göre $DW=1,617$ dir. $n = 16$, $k = 1$ ve $\alpha = 0,05$ iken DW tablosundan $D_L = 1,106$ ve $D_U = 1,371$ bulunmuştur. Buna göre $4 - D_L = 4 - 1,106 = 2,894$ ve $4 - D_U = 4 - 1,371 = 2,629$ olup, $1,371 < 1,617 < 2,629$, yani $D_U < DW < 4 - D_U$ olduğundan $H_0: \rho = 0$ hipotezi kabul edilir ve otokorelasyon yoktur. Buna göre ardışık hata terimleri arasındaki ilişki önemsizdir.

f) Çoklu lineer regresyon modeli: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$

Tahmin denklemi: $\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4$

$\hat{Y} = 53,292 + 0,161 X_1 + 0,170 X_2 + 0,333 X_3 + 0,521 X_4$

Tablo 7.11 Değişkenlere Ait Korelasyon Katsayıları

		SKB	Yaş	Boy	Ağırlık	GİSS
SKB	Pearson Correlation	1	,285	,725**	,805**	,914**
	Sig. (2-tailed)		,284	,001	,000	,000
	N	16	16	16	16	16
Yaş	Pearson Correlation	,285	1	-,251	-,090	,308
	Sig. (2-tailed)	,284		,349	,739	,246
	N	16	16	16	16	16
Boy	Pearson Correlation	,725**	-,251	1	,894**	,685**
	Sig. (2-tailed)	,001	,349		,000	,003
	N	16	16	16	16	16
Ağırlık	Pearson Correlation	,805**	-,090	,894**	1	,739**
	Sig. (2-tailed)	,000	,739	,000		,001
	N	16	16	16	16	16
GİSS	Pearson Correlation	,914**	,308	,685**	,739**	1
	Sig. (2-tailed)	,000	,246	,003	,001	
	N	16	16	16	16	16

** . Correlation is significant at the 0.01 level (2-tailed).

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson	$n = 16, k = 4$ $\alpha = 0,05$
1	,942 ^a	,887	,846	4,79424	1,393	$D_L = 0,734,$ $4 - D_L = 3,266$ $D_U = 1,935$ $4 - D_U = 2,065$

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	1990,918	4	497,729	21,655	,000 ^b
Residual	252,832	11	22,985		
Total	2243,750	15			

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	53,292	65,883		,809	,436	-91,716	198,300		
	Yas	,161	,139	,166	1,162	,270	-,144	,467	,503	1,990
	Boy (*)	,170	,448	,100	,379	,712	-,817	1,156	,146	6,829
	Agirlik	,333	,260	,315	1,281	,226	-,239	,906	,169	5,902
	GISS	,521	,182	,562	2,872	,015	,122	,921	,268	3,737

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson	$n = 16, k = 4$ $\alpha = 0,05$
1	,941 ^a	,886	,857	4,61995	1,388	$D_L = 0,734,$ $4 - D_L = 3,266$ $D_U = 1,935$ $4 - D_U = 2,065$

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	1987,623	3	497,729	21,655	,000 ^b
Residual	256,127	12	22,985		
Total	2243,750	15			

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	77,554	14,753	5,257	,000	45,409	109,698		
	Yaş	,134	,115	,138	,265	-,116	,385	,683	1,464
	Ağırlık	,403	,176	,381	,041	,019	,787	,343	2,916
	GİSS	,548	,162	,590	,005	,195	,900	,313	3,195

a. Dependent Variable: SKB