

ÇOKLU REGRESYON

Regresyon analizinin amaçları şunlardır:

- Bağımlı değişkeni etkilediği bilinen değişkenler yardımıyla bağımlı değişkeni kestirmek
- Bağımlı değişkeni etkilediği düşünülen bağımsız değişkenlerden hangisi ya da hangilerinin bağımlı değişkeni daha çok etkilediğini bulmak ve aralarındaki karmaşık yapıyı tanımlamak.
- Bağımsız değişken yada değişkenlerin bağımlı değişkende gözlenen değişimlerin ne kadarını açıkladığını determinasyon katsayısı ile belirlemek.

Regresyon analizinin varsayımları:

- 1) Örneklem büyüklüğü
- 2) Kayıp değerler
- 3) Aykırı değerler
- 4) Çoklu bağıntı ve teklik
- 5) Normallik
- 6) Doğrusallık ve Artıkların eş varyanslılığı
- 7) Hataların bağımsızlığı

Araştırma problemi: Uluslararası öğrenci değerlendirme projesi PISA 2006 sonuçlarına göre Türkiye’de ailelerin bazı soyso-ekonomik indeksleri öğrencilerin fen başarısını nasıl etkilemektedir?

Bu alt problemde ailelerin soyso-ekonomik durumlarını gösteren bazı indeksler kullanılarak, sosyo-ekonomik özelliklerin fen başarısı üzerinde etkisi belirlenmeye çalışılmıştır.

Araştırma kapsamındaki değişkenler incelendiğinde ise:

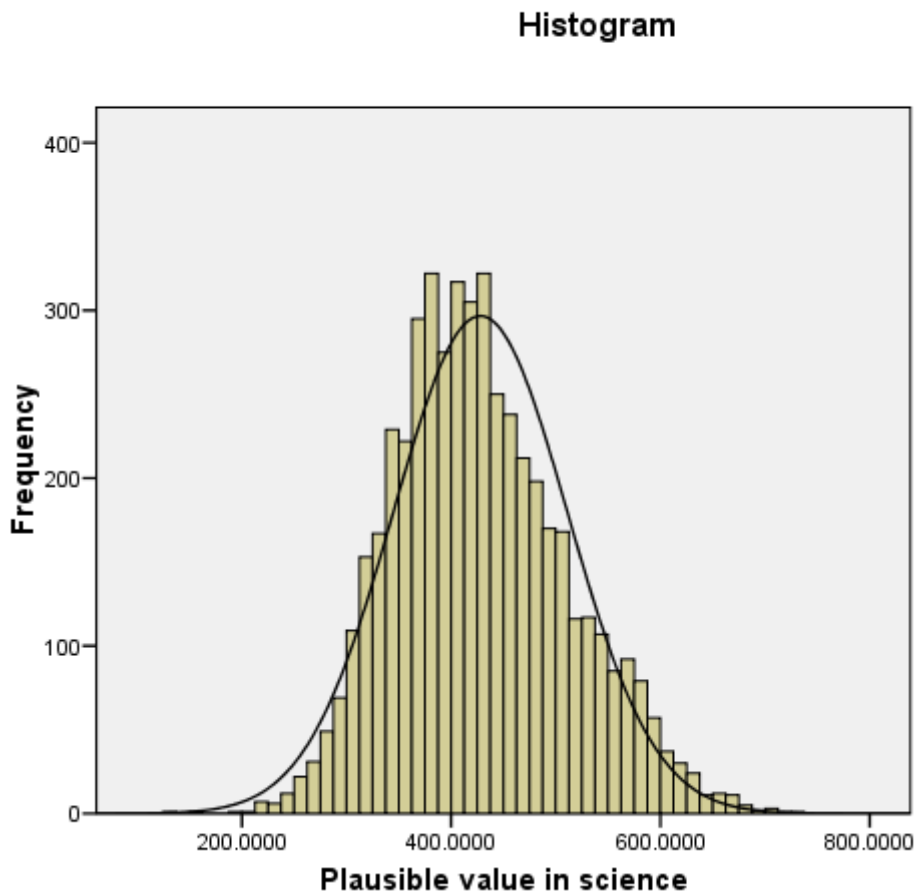
- Bağımlı değişken:
 - Öğrencilerin fen başarısı
- Bağımsız değişkenler
 - Evdeki kültürel varlıklar indeksi
 - Evdeki eğitimsel kaynaklar indeksi

- Evdeki varlıklar indeksi
- Aile zenginlik indeksi
- Sosyo-ekonomik ve kültürel indeks

VERİLERİN İNCELENMESİ VE VARSAYIMLARIN TEST EDİLMESİ

Örneklem büyüklüğü

Veride 4942 kişi yer almaktadır. Bağımlı değişken olan fen başarısına ilişkin puanların (plausible value in science) dağılımı incelendiğinde dağılımın normallik varsayımını sağladığı görülmektedir.



Örneklem büyüklüğü çoklu korelasyon testi için $N \geq 50 + 8m$ yordayıcıların testi için $N \geq 104 + m$ alınabilir (m: bağımsız değişken sayısı) veri sayısı bu koşulları sağlamaktadır.

Kayıp Değerler

Geniş veri setlerinde seçkisiz bir örüntü sergileyen az sayıda kayıp değer varsa, sorun çok ciddi değildir ve kayıp değerleri ortadan kaldırmada farklı yöntemlerin kullanılması benzer sonuçlar üretecektir. Ancak küçük ya da orta büyüklükteki veri setlerinde çok sayıda kayıp değer bulunması ciddi sorunlara neden olacaktır (Tabachnick & Fidell, 2007). Bu sebeple öncelikle veri setinde kayıp veri olup olmadığı incelenecektir.

Tablo 1. Kayıp değerler

	Evdeki kültürel varlıklar indeksi	Evdeki eğitimsel kaynaklar indeksi	Evdeki varlıklar indeksi	Aile zenginlik indeksi	Sosyo-ekonomik ve kültürel indeks
N Geçerli değer	4913	4934	4941	4941	4934
Kayıp değer	29	8	1	1	8

Tablo1 incelendiğinde kayıp değerler sayıca oldukça düşük çıkmıştır. Başlangıçta 4942 olan örneklem büyüklüğü, eksik verili bireylerin veri setinden çıkarılması ile 4905'e düşmüştür. Yani eksik verinin çıkarılması, örneklem büyüklüğünü oldukça az etkilemiştir.

Aykırı Değerler

Alışılageldik değerlerin dışında değerlere ya da aşırı değerlere sahip olan (dağılımın uçlarında yer alan) denekler uç değerler olarak adlandırılır (Çokluk, Şekercioğlu, & Büyüköztürk, 2010). Uç değerler iki farklı şekilde ortaya çıkar: tek yönlü uç değer ve çok yönlü uç değer

Tek yönlü aykırı değerlerin belirlenmesi için standartlaştırılmış puanlara bakılır. $z \text{ değeri} > |3|$ olanlar aykırı değerdir.

Tablo 2. Z puanları

	Evdeki kültürel varlıklar indeksi	Evdeki eğitimsel kaynaklar indeksi	Evdeki varlıklar indeksi	Aile zenginlik indeksi	Sosyo-ekonomik ve kültürel indeks
N	4905	4905	4905	4905	4905
Minimum	-1.65141	-2.81269	-4.17346	-2.35842	-2.93903
Maximum	1.34932	1.55021	3.38654	3.77411	3.08037

Tablo 2 incelendiğinde evdeki varlıklar indeksi ve sosyo ekonomik kültürel indeks değişkenlerinde z puanlarının +3 -3 aralığının sınırlarını geçtiği görülmektedir. Bu denekler veriden çıkartılmadan önce çok değişkenli uç değerler incelenmelidir.

Çok değişkenli aykırı değerler regresyon içerisindeki Mahalanobis uzaklıkları ile incelenir. Bu işlem output'ta bir çıktı vermez ancak veri dosyasında ekstra bir değişken olarak karşımıza çıkar (Mah_1). Çok yönlü uç değer olduğunu belirlemek için kritik ki-kare değerinin belirlenmesi gerekmektedir. Çok yönlü uç değerler için kabul edilen ölçüt, $p < .001$ düzeyide manidar Mahalanobis uzaklığı değeridir.

Değişken sayımıza göre belirlediğimiz kritik ki-kare değeri 20,517'dir. Analiz sonucu elde edilen en büyük Mahalanobis uzaklığı değeri, kritik değerden büyüktür ($40.67 > 16.27$). Bu duruma uyan 32 veri vardır. Bu veriler çıkartıldıktan sonra 4873 veri kalmıştır. Bu verilerde tek yönlü uç değer için tekrar incelenmiştir.

Tablo 3. Çok yönlü uçdeğerler çıkartıldıktan sonra Z puanları

	Evdeki kültürel varlıklar indeksi	Evdeki eğitimsel kaynaklar indeksi	Evdeki varlıklar indeksi	Aile zenginlik indeksi	Sosyo-ekonomik ve kültürel indeks
N	4873	4873	4873	4873	4873
Minimum	-1.65141	-2.81269	-3.14752	-2.35842	-2.51916
Maximum	1.34932	1.55021	3.38654	3.77411	3.08037

Tablo 4. Çok yönlü uçdeğerler çıkartıldıktan sonra betimsel istatistikler

		Evdeki kültürel varlıklar indeksi	Evdeki eğitimsel kaynaklar indeksi	Evdeki varlıklar indeksi	Aile zenginlik indeksi	Sosyo-ekonomik ve kültürel indeks
N	Valid	4873	4873	4873	4873	4873
Ortalama		.00434	-.63366	-1.03939	-1.48524	-1.23008
Ortanca		.1844	-.9119	-.9579	-1.6157	-1.3566
Mod		.1844	-.9119	-.9579	-1.6157	-2.2971
Çarpıklık katsayısı		-.157	.205	-.103	-.090	.388
Çarpıklık katsayısının standart hatası		.035	.035	.035	.035	.035
Basıklık katsayısı		-.940	-.856	-.032	-.133	-.388
Basıklık katsayısının standart hatası		.070	.070	.070	.070	.070

Normallik

Çok değişkenli istatistikler de normallik varsayımı tek değişkenli ve çok değişkenli olarak ayrı ayrı incelenir. Tek değişkenli normallik, örnekleme bir değişkene ilişkin gözlemlerin normal dağılım şekli gösterdiği anlamına gelir (Çokluk, Şekercioğlu, & Büyüköztürk, 2010). Tek değişkenli normalliği değerlendirmek amacıyla her değişken için histogram ve normal Q-Q grafikleri çizilebileceği gibi çarpıklık-basıklık katsayılarının incelenmesi ve Kolmogorov-Smirnov hipotez testinin değerlendirilmesi diğer alternatiflerdir.

Tek değişkenli normalliği belirlemenin bir diğer yolu olan çarpıklık-basıklık katsayıları ile ilgili elde edilen değerler aşağıdaki tabloda sunulmuştur. Standart normal dağılımda basıklık ve çarpıklık katsayıları sıfırdır. Bu değerlerin (-1, +1) arasında kalması, dağılımın normalden aşırı sapma göstermediğinin bir kanıtı olarak değerlendirilebilir.

Tablo 5.Çarpıklık ve basıklık katsayıları

	Çarpıklık		Basıklık	
	Değer	Std. hata	Değer	Std. hata
Evdeki kültürel varlıklar indeksi	-.157	.035	-.940	.070
Evdeki eğitimsel kaynaklar indeksi	.205	.035	-.856	.070
Evdeki varlıklar indeksi	-.103	.035	-.032	.070
Aile zenginlik indeksi	-.090	.035	-.133	.070
Sosyo-ekonomik ve kültürel indeks	.388	.035	-.388	.070

Çoklu Bağlantı ve Tekillik

Bağımsız değişkenler arasındaki korelasyon oldukça yüksek olduğunda, çoklu bağlantı problemi ile karşılaşmaktadır. Özellikle korelasyon değerinin 0.80 veya 0.90'dan büyük çıkması durumunda dikkat edilmelidir.

Tablo 6. Değişkenler Arası Korelasyonlar

	Evdeki kültürel varlıklar indeksi	Evdeki eğitimsel kaynaklar indeksi	Evdeki varlıklar indeksi	Aile zenginlik indeksi	Sosyo-ekonomik ve kültürel indeks
Evdeki kültürel varlıklar indeksi	1				
Evdeki eğitimsel kaynaklar indeksi	.322	1			
Evdeki varlıklar indeksi	.563	.787	1		
Aile zenginlik indeksi	.310	.672	.918	1	
Sosyo-ekonomik ve kültürel indeks	.444	.632	.803	.733	1

Tablo 6 incelendiğinde bağımsız değişkenlerin birbirleriyle olan ilişkilerinin bazılarının 0.70’den büyük çıktığı görülmektedir.

Çoklu bağlantı, ayrıca “Variance Inflation Factor, VIF” ile kontrol edilir. VIF değerinin 10’dan büyük olması çoklu bağlantı problemine işaret etmektedir (Cohen & Cohen, 1983). Ayrıca analizler sonucu elde tolerans değeri de çoklu bağlantı ile ilgili bilgi vermektedir. Tolerans değerinin 0.10’dan büyük olması çoklu bağlantı probleminin olmadığına işaret etmektedir (Field, 2005; Akt. Çokluk, Şekercioğlu, & Büyüköztürk, 2010).

Bu araştırmadan elde edilen VIF ve tolerans değerleri aşağıdaki tabloda verilmektedir. 10’dan büyük VIF değerine rastlandığı ve tolerans değerlerinin de 0.10’dan küçük olan iki değişken olduğu görülmektedir.

Tablo 7. VIF ve Tolerans Değerleri

	Tolerans	VIF
Evdeki kültürel varlıklar indeksi	.311	3.215
Evdeki eğitimsel kaynaklar indeksi	.276	3.625
Evdeki varlıklar indeksi	.034	29.740
Aile zenginlik indeksi	.072	13.863
Sosyo-ekonomik ve kültürel indeks	.355	2.821

İnceleme sonunda bağımsız değişkenler arasında çoklu bağlantı görülmüştür. En yüksek VIF değeri en düşük tolerans değeri olan evdeki varlıklar indeksi değişkeninin analizin dışında tutulmasına karar verilmiştir.

Tablo 8.Evdeki varlıklar indeksi çıktıktan sonra VIF ve Tolerans Değerleri

	Tolerans	VIF
Evdeki kültürel varlıklar indeksi	.798	1.254
Evdeki eğitimsel kaynaklar indeksi	.503	1.986
Aile zenginlik indeksi	.389	2.571
Sosyo-ekonomik ve kültürel indeks	.386	2.587

Evdeki varlıklar indeksi analiz dışı yapıldıktan sonra 10'dan büyük VIF değerine rastlanmamıştır. Belirtilen değişken çıkartıldıktan sonra bağımsız değişkenler arası korelasyon katsayısı aşağıdaki tabloda verilmiştir.

Tablo 7. Değişkenler Arası Korelasyonlar

	Evdeki kültürel varlıklar indeksi	Evdeki eğitimsel kaynaklar indeksi	Aile zenginlik indeksi	Sosyo-ekonomik ve kültürel indeks
Evdeki kültürel varlıklar indeksi	1			
Evdeki eğitimsel kaynaklar indeksi	.322	1		
Aile zenginlik indeksi	.310	.672	1	
Sosyo-ekonomik ve kültürel indeks	.444	.632	.733	1

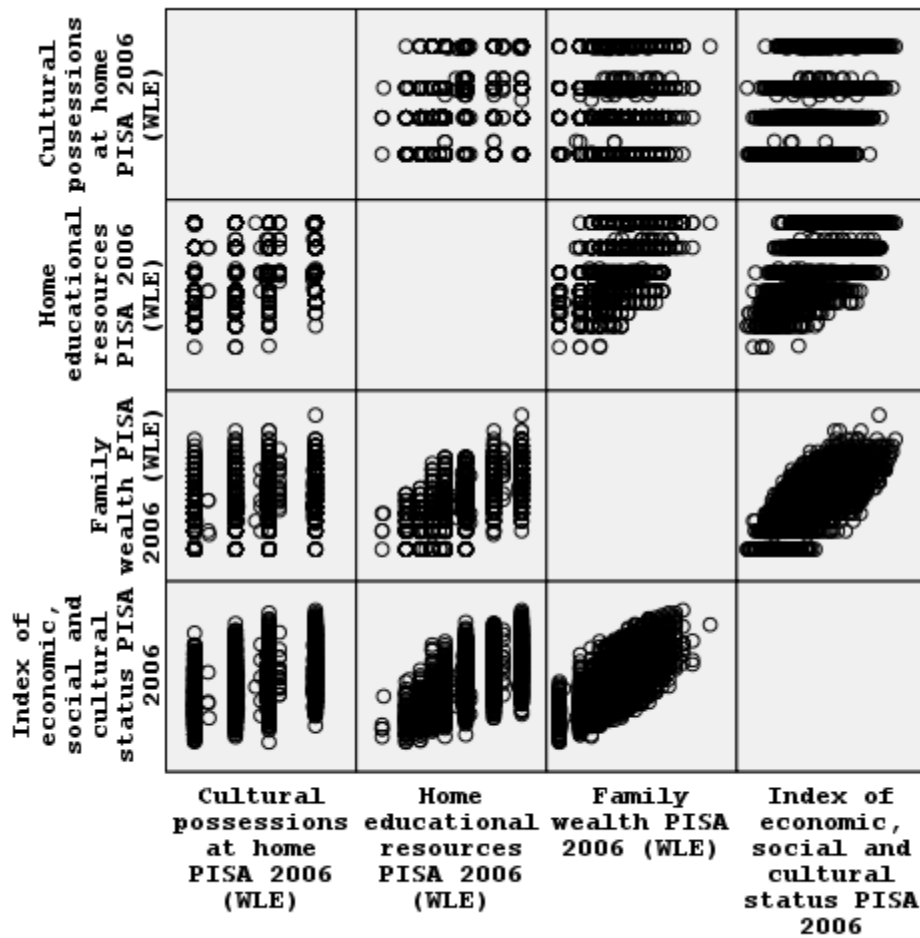
Doğrusallık ve Artıkların eşvaryanslılığı

Artıkların (gözlenen ve kestirilen bağımlı değişken değerleri arasındaki fark) saçılım grafiklerinin incelenmesi, normallik, doğrusallık ve kestirilen bağımlı değişken ve kestirilen hata arasındaki eş varyanslılık varsayımlarının testini sağlar.

Eş varyanslılık kestirim hatalarının standart sapmasının kestirilen tüm bağımlı değişken değerleri için eşit olmasıdır. Eş varyanslılığın bozulması bağımsız değişkenlerden bazılarının çarpık olmasından kaynaklanabilir. Bu durumda dönüşüm yapılabilir.

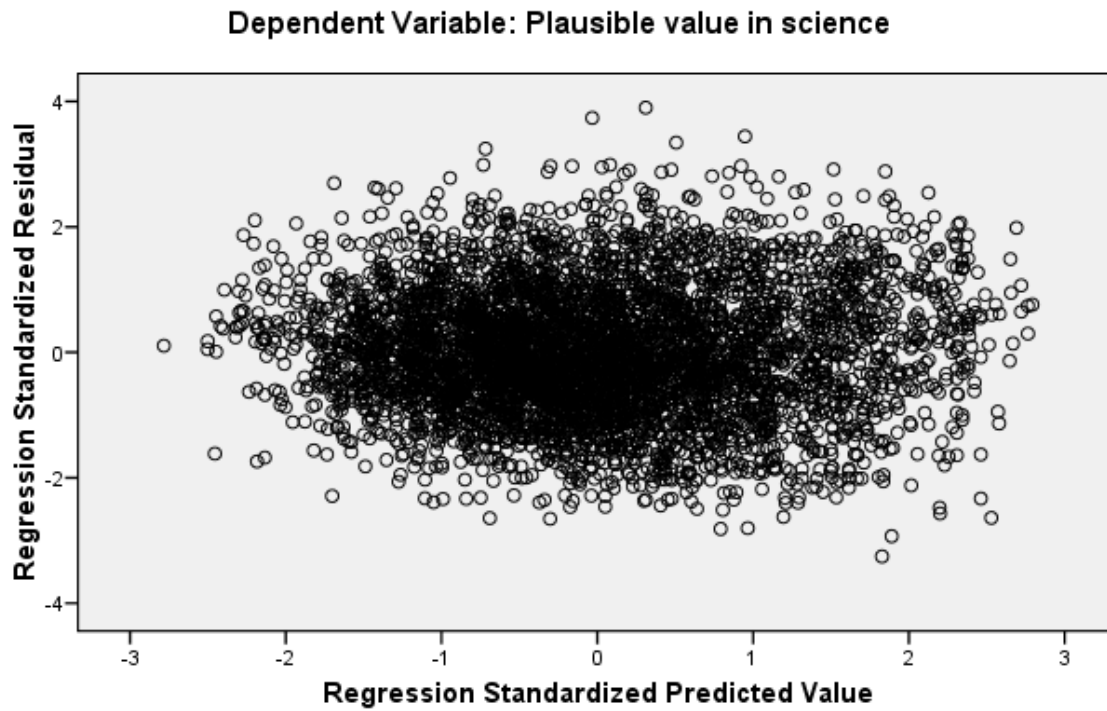
Eş varyanslılık, bağımsız değişkenlerden birinin denkleme alınmayan başka bir bağımsız değişkenle etkileşiminden de kaynaklanabilir. Artıkların saçılım grafiklerinin incelenmesi, normallik, doğrusallık ve kestirilen bağımlı değişken ve kestirilen hata arasındaki eş varyanslılık varsayımlarının testini sağlar.

Doğrusallığın ve eş varyanslılığın bozulması regresyon katsayılarını geçersiz kılmaz ama etkisini zayıflatır. Değişkenler arasındaki doğrusal olmayan ilişkiler, doğrusal korelasyon katsayıları ile ortaya açığa çıkarılamaz.



Saçılım grafiklerinin elips şeklinde olması çok değişkenli normalliğin ve doğrusallığın sağlandığını gösterir. Saçılım grafikleri incelendiğinde evdeki kültürel varlıklar değişkeninde normallikten sapmalar olduğu görülmüştür. Belirtilen değişkende dönüşümler yapılmasına rağmen olumlu bir değişiklik olmamıştır.

Doğrusallığın incelenmesinde diğer bir yol ise artık grafiklerini incelemektir. Regresyon analizindeki artık grafiği ile incelenir, eğer noktalar eğrisel bir görünüm oluşturduysa, artıklar bazı yordanan değerler için grafiğin altında bazısı için ise üstünde yer alıyorsa doğrusallık bozulur. Eğer noktalar sıfır çizgisi etrafında kümelenmiş ise doğrusallık karşılanıyor demektir.



Artık grafiği incelendiğinde doğrusallığın sağlandığı sonucuna ulaşılmıştır.

Bu yapılan analiz sonunda ise evdeki kültürel varlıklar değişkeninin de analize dahi edilmemesine karar verilmiştir.

Hataların Bağımsızlığı

Otokorelasyon, çoklu regresyon analizinde hata teriminin birbirini izleyen değerleri arasında ilişki bulunması halidir. Bu durum, genel doğrusal regresyon modelinin önemli bir varsayımından sapmadır. Genel doğrusal regresyon modeli varsayım gereği olarak, hata terimleri arasında bir ilişki yoktur.

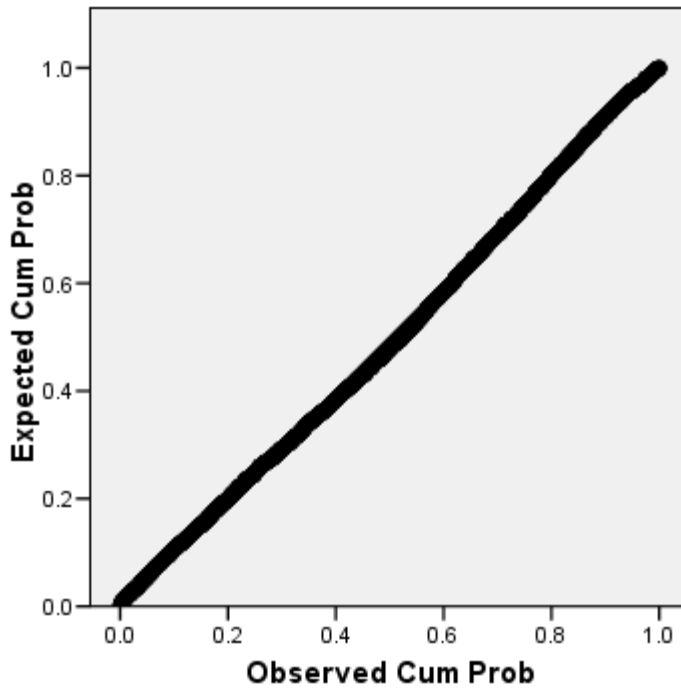
Durbin-Watson testi ise otokorelasyonu* test etmek için kullanılmaktadır. SPSS'te Regresyon analizi ile hesaplanmaktadır (Statistics'te Durbin-Watson işaretlendiği zaman elde edilebilir).

Durbin Watson 0 ile 4 arası değer almaktadır. 1.5 ve 2.5 arasında olmalıdır, bu durum gözlemlerin bağımsızlığını göstermektedir.

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.429(a)	.184	.183	76.6314919	2.057

Dependent Variable: Plausible value in science



ÇIKTILARIN YORUMLANMASI

ÇOKLU REGRESYON

Tablo. Fen başarısını etkileyen sosyo-ekonomik değişkenlere yönelik regresyon analizi sonuçları

model	Standardize edilmemiş		Standardize edilmiş	t	Sig.	F	p
	B	Standart hata	Regresyon katsayısı (β)				
sabit	452.287	2.100		215.333	.000	335.834	.000(a)

Evdeki eğitimsel kaynaklar	11.815	1.174	.184	10.061	.000
Aile zenginlik indeksi	-14.829	1.738	-.178	-8.532	.000
Sosyo- ekonomik ve kültürel indeks	30.875	1.536	.401	20.096	.000
R =,414 R ² = .171 adjusted R ² = .171					

Regresyon analizi sonuçlarına göre, evdeki eğitimsel kaynaklar, aile zenginlik indeksi, sosyo-ekonomik ve kültürel indeks değişkenleri birlikte, fen başarı puanları ile düşük düzeyde anlamlı bir ilişki vermektedir (R =,414 R²= .171) bağımsız değişkenler fen başarısındaki toplam varyansın yaklaşık olarak %17.1'ini açıklamaktadır.

Standardize edilmiş regeryon katasyılarına göre yordayıcı değişkenlerin fen başarısında görelî önem sırası; soyso-ekonomik ve kültürel indeks, evdeki eğitimsel kaynaklar ve aile zenginlik indeksidir. T puanları incelendiğinde 3 değişkeninde anlamlı birer yordayıcı oldukları ortaya çıkmıştır.

Regresyon analizi sonuçlarına göre regresyon denklemi:

Fen başarısı: $452.287 + 11.815 * (\text{evdeki eğitimsel kaynaklar}) + (-14.829) * (\text{aile zenginlik indeksi}) + (30,875) * (\text{sosyo-ekonomik ve kültürel indeks})$

Buna göre aile zenginlik indeksinin fen başarı puanı ile negatif yönlü bir etkisinin olduğu görülmektedir. Evdeki eğitimsel kaynaklar ve sosyo-ekonomik ve kültürel indekslerinin ise fen başarısında ile pozitif etkisi olduğu görülmektedir. Buna göre ailelerin evdeki kültürel varlıklar indeksi ve sosyo-ekonomik kültürel indeksleri arttıkça öğrencilerin fen başarılarında olumlu etkileneceği söylenebilir.