

1996-1998 YILLARI ARASINDA GÜNEY CAROLINA'DA OKUL İÇİ ŞİDDET İSTATİSKLERİ ANALİZİ

(Bir Önceki Projeden Devam Edilecektir)

Hazırlayan : Süleyman Öğrekçi

No :29054041

1996 ve 1998 yılları arasında Güney Carolina'da resmi kayıtlara geçen okul içi şiddet olaylarını ele alan bir çalışma, *South Carolina Department of Public Safety Office of Justice Programs* kurumu tarafından hazırlanmış ve kamuoyuyla paylaşılmıştır. Yapılan çalışmanın sonuç belgesi (final report) http://www.jrsa.org/ibrrc/background-status/South_Carolina/sc_report.pdf adresinden ulaşılabilir durumdadır. hazırladığım bu projede bu çalışmanın 381 deneği üzerinden veriler toplanmış olup, bazı incelemeler yapılmış sonuç istatistikler paylaşılmıştır. Bu çalışmada toplanan istatistiklerin bir kısmı yer almaktadır. Örneğin tüm denekler yerine 381 denek, tüm değişkenler yerine belli sayıda değişken (örneğin suçun hangi ayda işlendiğine dair değişken gibi bazı değişkenler yoktur) kullanılmıştır.

Proje hakkında kısa bir bilgi vermem gerekirse, projede 381 denek 8 değişken üzerinden değerlendirilmiştir. Aşağıda değişkenlerin detaylarını kabaca görebileceğimiz görüntü vardır.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	suc	String	3	0	Suç	{09A, Cinay...	None	4	Left	Nominal
2	yas	Numeric	9	0	Yaş	None	None	8	Right	Scale
3	etnik	String	1	0	Etnik Kökeni	{H, İspanyol...	None	6	Left	Nominal
4	irk	String	1	0	İrki	{A, Asyalı}...	None	5	Left	Nominal
5	cins	String	1	0	Cinsiyeti	{F, Bayan}...	None	4	Left	Nominal
6	silah	Numeric	9	0	Kullandığı Silah	{1, Silahsız}...	None	9	Right	Scale
7	yer	Numeric	9	0	Yakalandığı Yer	{2, Bar/Gec...	None	8	Right	Scale
8	yil	Numeric	8	0	Yakalanma Yılı	None	None	8	Right	Scale
9										

Deneklere ilişkin bilgilerin de bir örnek görüntüsü aşağıda verilmiştir.

1.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : suc 13B

	suc	yas	etnik	irk	cins	silah	yer	yil
1	13B	17	N	W	M	95	22	1998
2	13A	16	N	B	F	95	22	1998
3	13C	15	N	B	F	95	22	1998
4	13C	13	N	B	F	95	22	1998
5	13C	34	N	W	M	95	22	1998
6	120	12	N	W	M	95	22	1998
7	13A	14	N	W	M	95	22	1998
8	13B	15	N	B	M	95	22	1998
9	13B	13	U	B	M	95	22	1998

Aşağıdaki verilen iki tablo ile de çalışma ile elde edilen istatistiklerin bir özeti bulunmaktadır. Tablolarda değişkenlerin ve durumların sayısı, ortalama değerleri, standart sapmaları, maksimum ve minimum değerleri, kurtosis ve skewness değerleri gibi bazı sayısal bilgiler verilmiştir.

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Yaş	381	66	8	74	16,08	,351	6,849	46,907	5,026	,125	30,302	,249
Kullandığı Silah	381	0	95	95	95,00	,000	,000	,000				
Yakalandığı Yer	381	9	13	22	21,97	,026	,504	,254	-16,174	,125	274,797	,249
Yakalanma Yılı	381	2	1996	1998	1997,11	,041	,802	,642	-,207	,125	-1,415	,249
Valid N (listwise)	381											

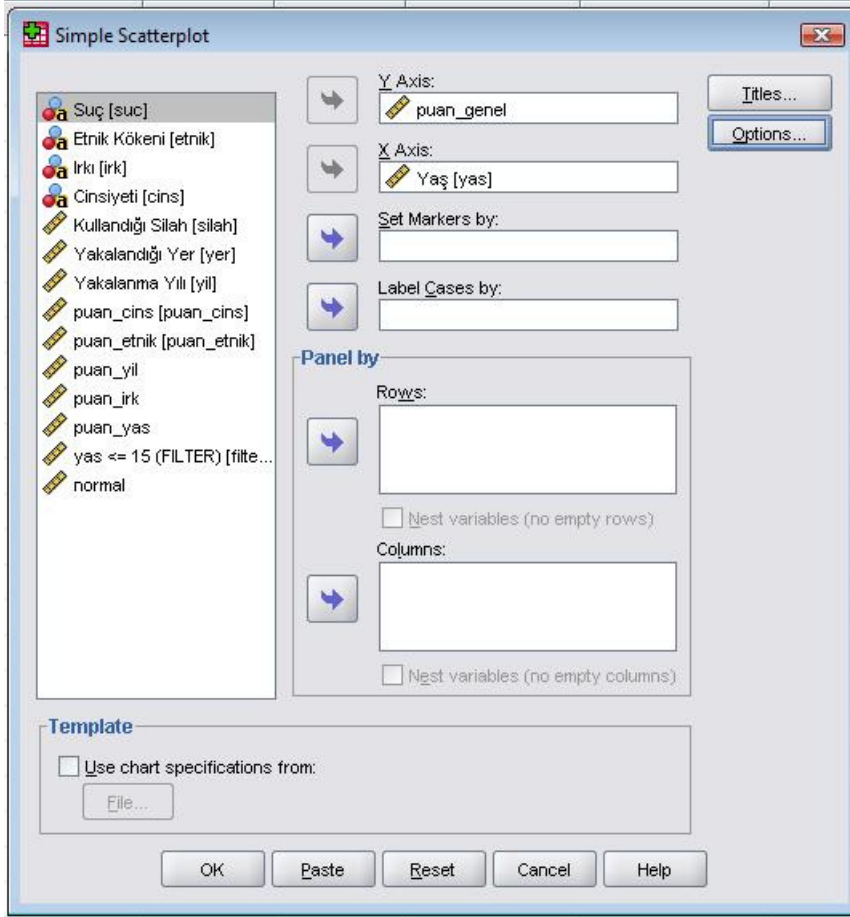
		Suç	Yaş	Etnik Kökeni	ırkı	Cinsiyeti	Kullandığı Silah	Yakalandığı Yer	Yakalanma Yılı
N	Valid	381	381	381	381	381	381	381	381
	Missing	0	0	0	0	0	0	0	0
Mean			16,08				95,00	21,97	1997,11
Std. Error of Mean			,351				,000	,026	,041
Median			15,00				95,00	22,00	1997,00
Mode			15				95	22	1998
Std. Deviation			6,849				,000	,504	,802
Variance			46,907				,000	,254	,642
Skewness			5,026					-16,174	-,207
Std. Error of Skewness			,125				,125	,125	,125
Kurtosis			30,302					274,797	-1,415
Std. Error of Kurtosis			,249				,249	,249	,249
Range			66				0	9	2
Minimum			8				95	13	1996
Maximum			74				95	22	1998
Sum			6125				36195	8369	760900

Yukarıdaki tablolara bakaram istatistik açıdan bazı yorumlar yapılabilir. Örneğin kurtosis değerleri dikkate alınarak yaş değişkeninin dağılım grafiğinin dik, yakalanma yılının dağılımının grafiğinin ise basık yapıda olduğu söylenebilir. Maksimum ve minimum değerleri dikkate alınarak da deneklerin yaşlarının 8 ile 74 arasında değiştiği söylenebilir.

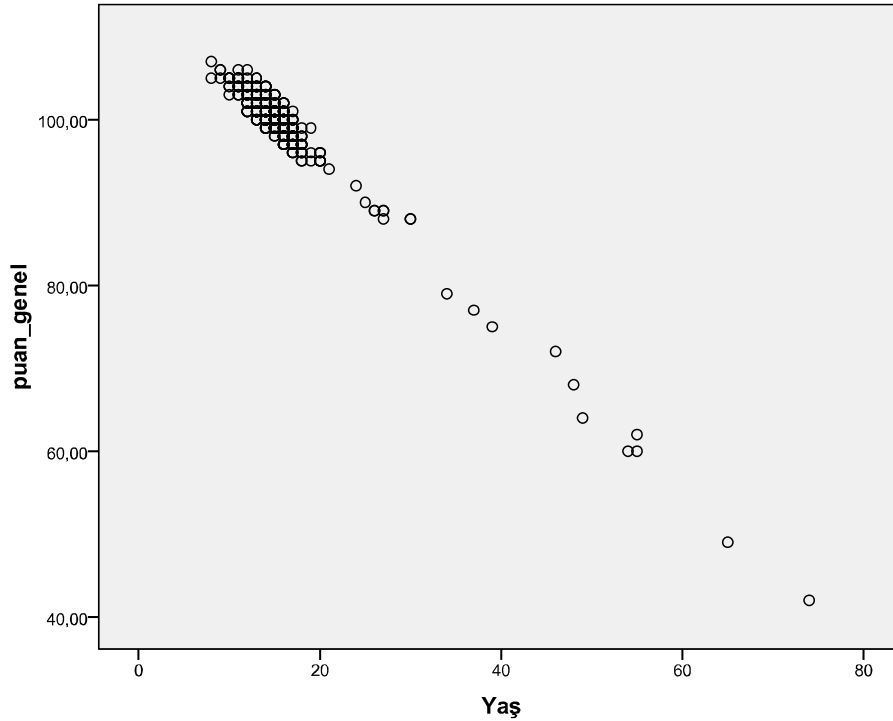
Şimdi SPSS ile korelasyon (correlation) analizinin nasıl yapılacağını inceleyelim. Korelasyon katsayısı değişkenler arasında bir ilişki olup olmadığı, ilişki varsa bu ilişkinin yönü ve büyüklüğü konusunda bize bilgi verir. Korelasyon için çeşitli testler vardır biz Pearson yöntemini kullanacağız. Bu yolla elde edeceğimiz korelasyon katsayısı -1 ile +1 arasında bir değer olacaktır. Bu katsayının işareti değişkenler arasındaki ilişkinin yönünü, katsayının mutlak değeri de bu ilişkinin büyüklüğünü verir. Örneğin değişkenlerden biri artarken diğeri azalıyorsa negatif korelasyon, biri artarken diğeri de artıyorsa pozitif korelasyon katsayısı elde ederiz. Değişkenler birbirine bağımlı ise korelasyon katsayısı aralarındaki bağımlılık ne kadar büyük ise korelasyon katsayısı sıfırdan o kadar uzaktır.

Korelasyon bilgisini çeşitli yollardan da elde edebiliriz. Örneğin nokta-saçılım grafiği bu konuda bize bilgi verebilir. İki değişkenin nokta saçılım grafiğinde noktalar geniş bir alana yayılmış şekilde ise bu iki değişken arasında bağımlılık söz konusu değildir demektir. Fakat noktalar bir doğru şeklinde ise bu değişkenler bağımlı demektir, bu durumda korelasyon katsayısı yüksek çıkacaktır. Bu doğrunun yönünden korelasyon katsayısının işareti de tahmin edilebilir. Doğrunun yönü aşağı ise negatif, yukarı ise (değişkenler arttıkça) pozitif korelasyon katsayısı elde edilecektir demektir. Eğer grafikte noktalar rastgele yayılmamış, fakat doğru şekli de oluşturmuyorsa, örenğin bir eğri grafiği şeklinde kümelenmiş ise bu durumda Pearson korelasyon testi uygulayamayız. Uygularsak testteki significance değeri bize korelasyon katsayısının anlamsız olduğunu söyleyecektir.

Şimdi bir örnek üzerinde korelasyon analizi yapalım. Datamızdaki deneklerin genel puanları ile yaşlarının korelasyon katsayısını inceleyelim. İlk önce nokta saçılım grafiğinde bu değişkenlerin dağılımını görüp bazı tahminlerde bulunalım. Daha sonra da Pearson yöntemi ile korelasyon katsayısını hesaplayıp tahminimizle karşılaştıralım.

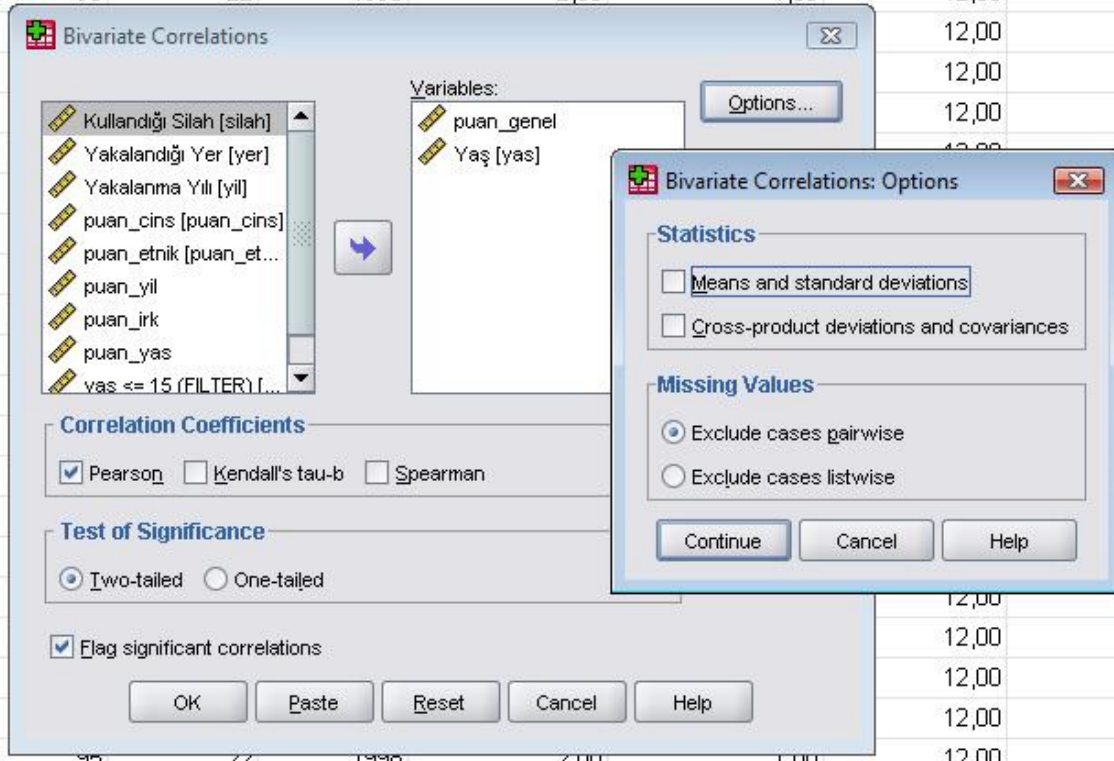


Yukarıdaki komutları verince aşağıdaki çıktıyı alırız.



Yukarıdaki grafiğe bakarak, noktalar bir doğru şekli çizdiği için puan ve yaş değişkenlerinin birbirine aşırı derecede bağımlı olduğunu ve büyük bir (1 e yakın) korelasyon katsayısı elde edeceğimizi öngörebiliriz. Ayrıca deneklerin yaşı arttıkça genel puanlarının azaldığını görüyoruz. Bundan dolayı korelasyon katsayısının negatif işaretli olacağı öngörülebilir. Sonuç olarak korelasyon katsayısı -1 e yakın bir değer olacaktır tahmininde bulunabiliriz.

Şimdi korelasyon katsayısını elde edelim. Bunun için Analyze/correlate/bivariate komutunu verip aşağıdaki işlemleri yapalım.



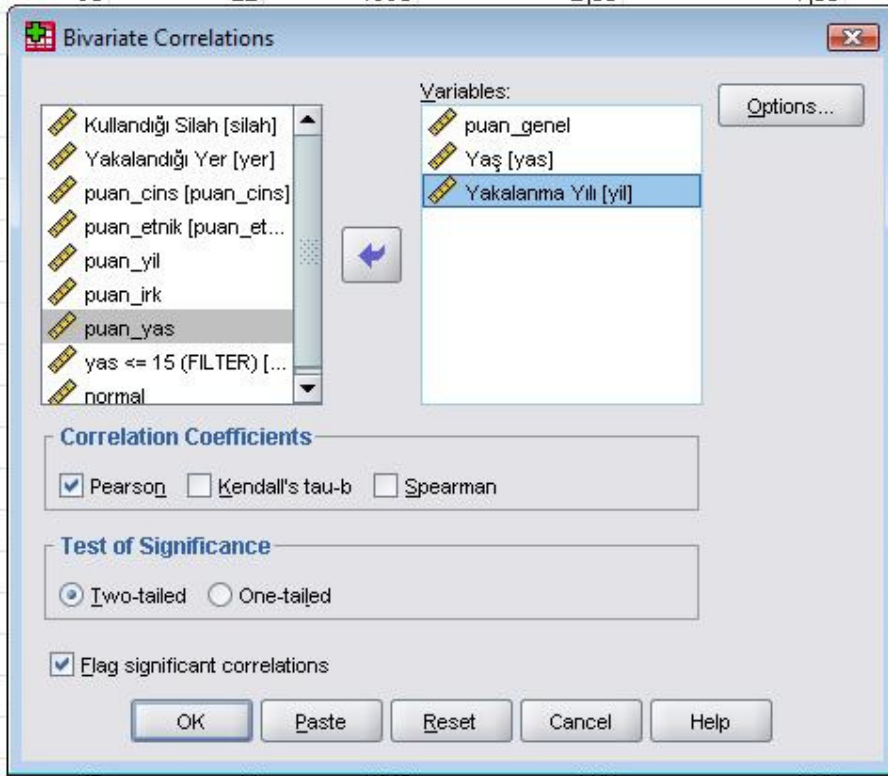
Yukarıdaki komutlar sonrasında aşağıdaki çıktı elde edilir.

Correlations			
		puan_genel	Yaş
puan_genel	Pearson Correlation	1	-,983**
	Sig. (2-tailed)		,000
	N	381	381
Yaş	Pearson Correlation	-,983**	1
	Sig. (2-tailed)	,000	
	N	381	381

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

Yukarıdaki tablodan da anlaşılacağı gibi puan ve yaş değişkenleri arasındaki korelasyon katsayısı -0.983... dür. Tahminlerimizdeki gibi -1 e yakın bir katsayı elde ettik.

Korelasyon analizi ikiden fazla değişken için de yapılabilir. Aşağıdaki örnekte bu gösterilmiştir.



Yukarıdaki komutlar sonrasında aşağıdaki çıktı alınır.

		puan_genel	Yaş	Yakalanma Yılı
puan_genel	Pearson Correlation	1	-,983**	-,105*
	Sig. (2-tailed)		,000	,040
	N	381	381	381
Yaş	Pearson Correlation	-,983**	1	-,005
	Sig. (2-tailed)	,000		,916
	N	381	381	381
Yakalanma Yılı	Pearson Correlation	-,105*	-,005	1
	Sig. (2-tailed)	,040	,916	
	N	381	381	381

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

* . Correlation is significant at the 0.05 level (2-tailed).

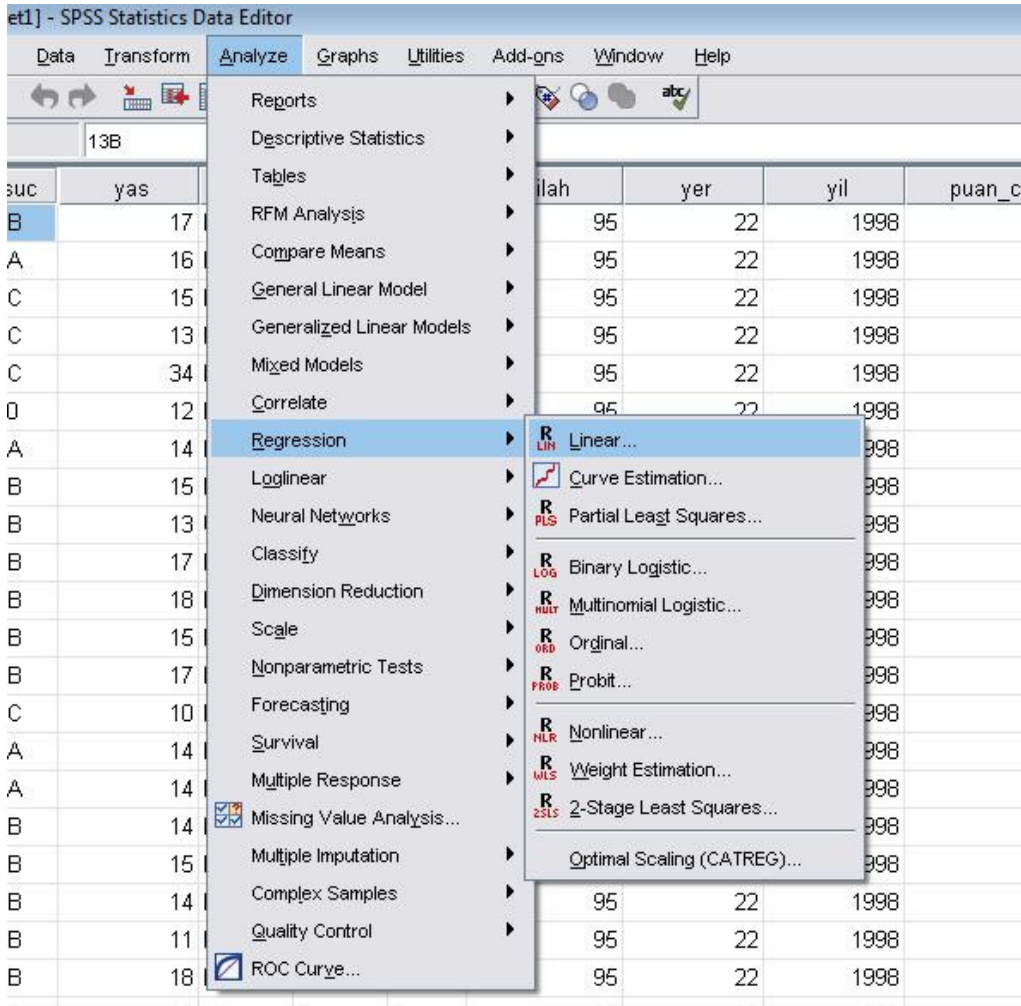
Yukarıdaki örnekte puan, yaş ve yakalanma yılı değişkenleri arasındaki korelasyon bilgisi elde edilmiştir. Tablodan özetle, puanla yaş arasında -0.983, puanla yakalanma yılı arasında -0.105, yaşla yakalanma yılı arasında -0,05 korelasyon katsayısı vardır.

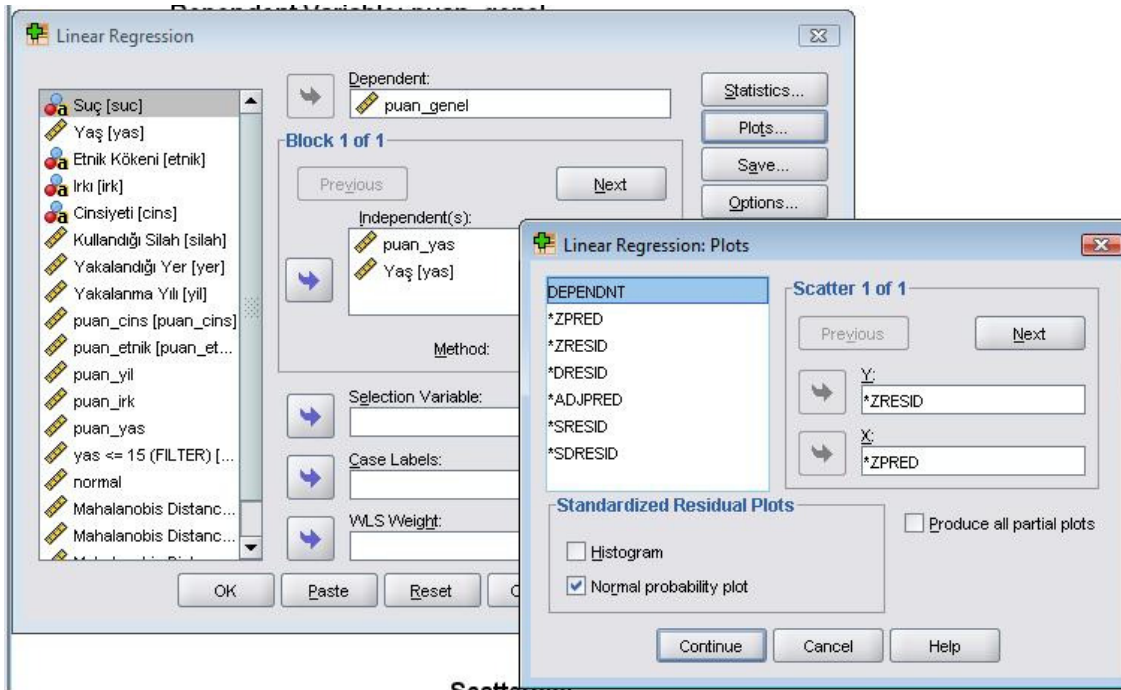
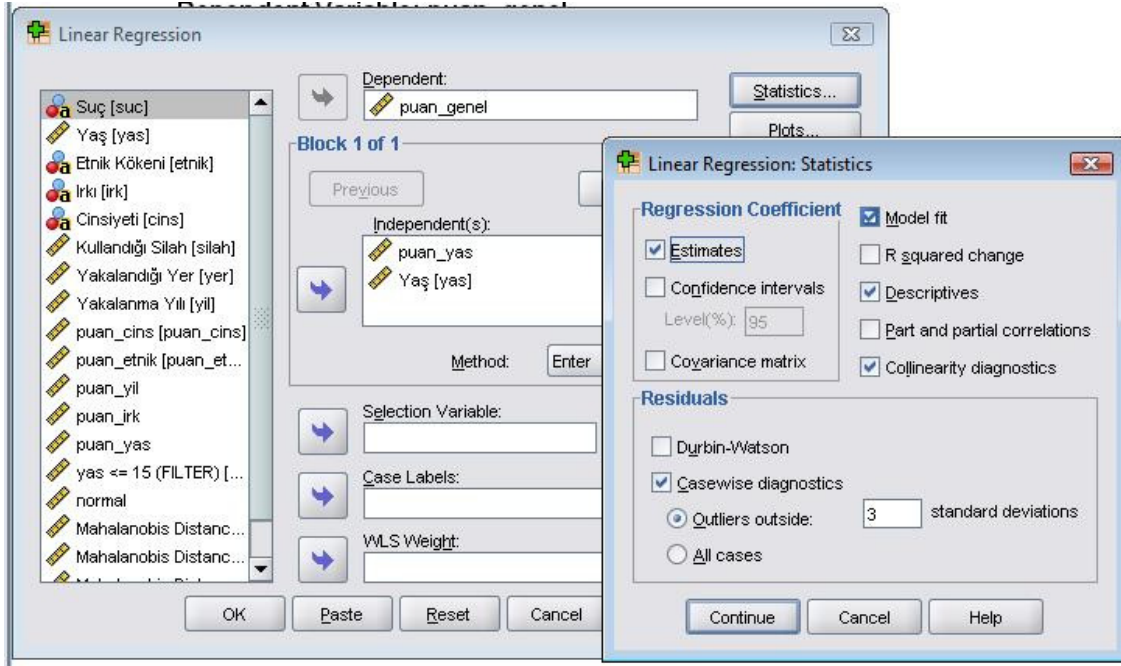
Şimdi iki veya daha fazla değişken arasında bir bağımlılık söz konusu mu, eğer bağımlılık varsa bu bağımlılığın kabaca özellikleri, büyüklüğü ve yönü hakkında, korelasyon analizi ile bilgi sahibi olabiliyoruz. Değişkenler arasındaki bağıntı hakkında daha fazla bilgi sahibi olmak için regresyon analizi yapılır.

Regresyon analizi ile değişkenler hakkında çok fazla bilgi edinebiliriz. Örneğin bir grup değişkenin sonucu hakkında tahminde bulunulabileceği ya da bulunulamayacağı konusunda bize bilgi verir. Ya da bir grup değişken içinden hangi değişkenin tahmin için en uygun olduğunu öğrenebiliriz. Veya tahmin edeceğimiz değişken başka bir değişkene bağlıysa yani, kontrol edilen başka bir değişkenden etkileniyorsa tahminimiz nasıl olmalıdır. Regresyon analizi bu gibi birçok konuda bize bilgi verir.

Şimdi bir örnek üzerinde regresyon analizini inceleyelim.

Aşağıdaki komutları sırasıyla verelim.





Yukarıdaki komutlar sonrasında aşağıdaki çıktıları alırız:

Descriptive Statistics

	Mean	Std. Deviation	N
puan_genel	99,0787	6,92053	381
puan_yas	83,9239	6,84889	381
Yaş	16,08	6,849	381

Correlations

		puan_genel	puan_yas	Yaş
Pearson Correlation	puan_genel	1,000	,983	-,983
	puan_yas	,983	1,000	-1,000
	Yaş	-,983	-1,000	1,000
Sig. (1-tailed)	puan_genel	.	,000	,000
	puan_yas	,000	.	,000
	Yaş	,000	,000	.
N	puan_genel	381	381	381
	puan_yas	381	381	381
	Yaş	381	381	381

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	Yaş ^a	.	Enter

a. Tolerance = ,000 limits reached.

b. Dependent Variable: puan_genel

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,983 ^a	,966	,965	1,28618

a. Predictors: (Constant), Yaş

b. Dependent Variable: puan_genel

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	17572,672	1	17572,672	10622,660	,000 ^a
	Residual	626,966	379	1,654		
	Total	18199,638	380			

a. Predictors: (Constant), Yaş

b. Dependent Variable: puan_genel

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	115,041	,168		683,520	,000		
	Yaş	-,993	,010	-,983	-103,066	,000	1,000	1,000

a. Dependent Variable: puan_genel

Excluded Variables^b

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	puan_yas	. ^a	.	.	.	,000	.	,000

a. Predictors in the Model: (Constant), Yaş

b. Dependent Variable: puan_genel

Collinearity Diagnostics^a

Model	Dimensi on	Eigenvalue	Condition Index	Variance Proportions	
				(Constant)	Yaş
1	1	1,920	1,000	,04	,04
	2	,080	4,905	,96	,96

a. Dependent Variable: puan_genel

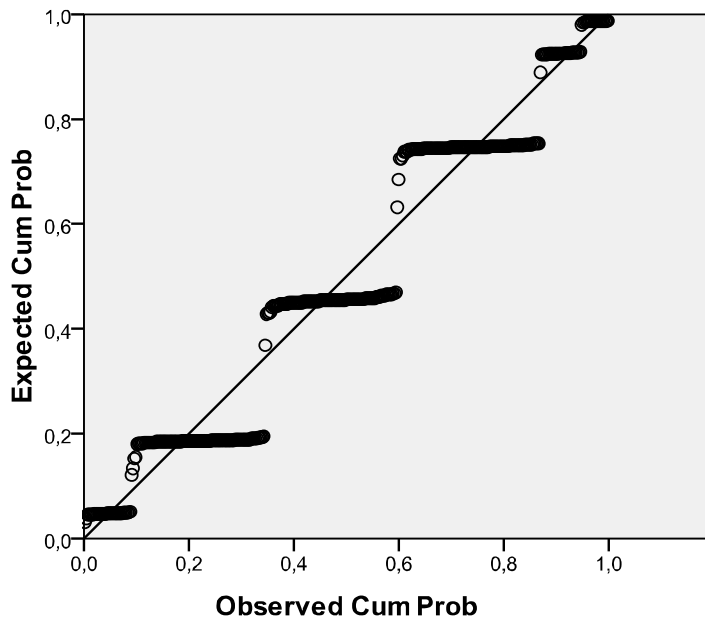
Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	41,5660	107,0975	99,0787	6,80028	381
Std. Predicted Value	-8,457	1,179	,000	1,000	381
Standard Error of Predicted Value	,066	,562	,079	,050	381
Adjusted Predicted Value	41,4636	107,1108	99,0793	6,79754	381
Residual	-2,38853	2,87407	,00000	1,28449	381
Std. Residual	-1,857	2,235	,000	,999	381
Stud. Residual	-1,919	2,239	,000	1,002	381
Deleted Residual	-2,55032	2,88433	-,00052	1,29240	381
Stud. Deleted Residual	-1,926	2,251	,000	1,004	381
Mahal. Distance	,000	71,528	,997	5,640	381
Cook's Distance	,000	,125	,003	,012	381
Centered Leverage Value	,000	,188	,003	,015	381

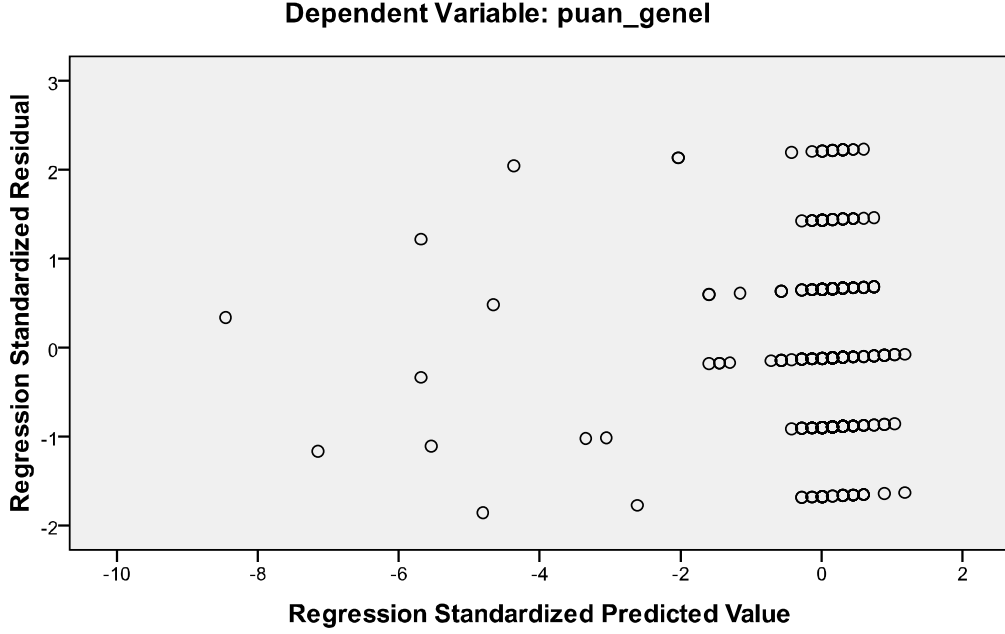
a. Dependent Variable: puan_genel

Normal P-P Plot of Regression Standardized Residual

Dependent Variable: puan_genel

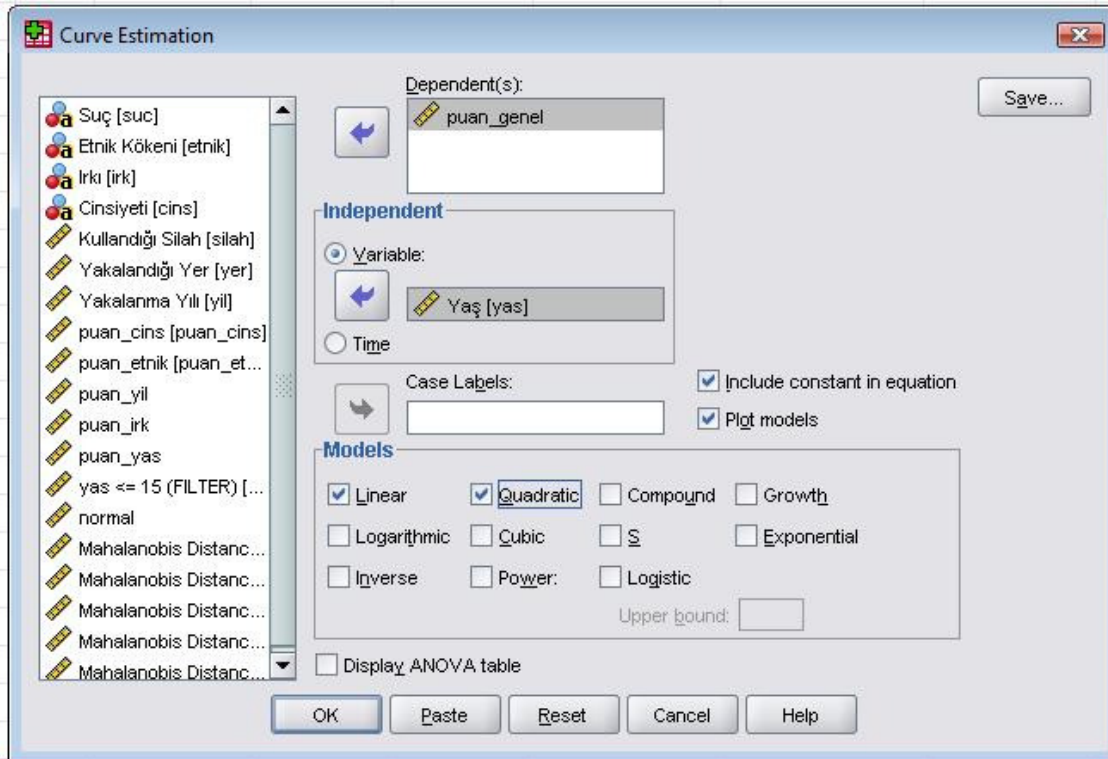
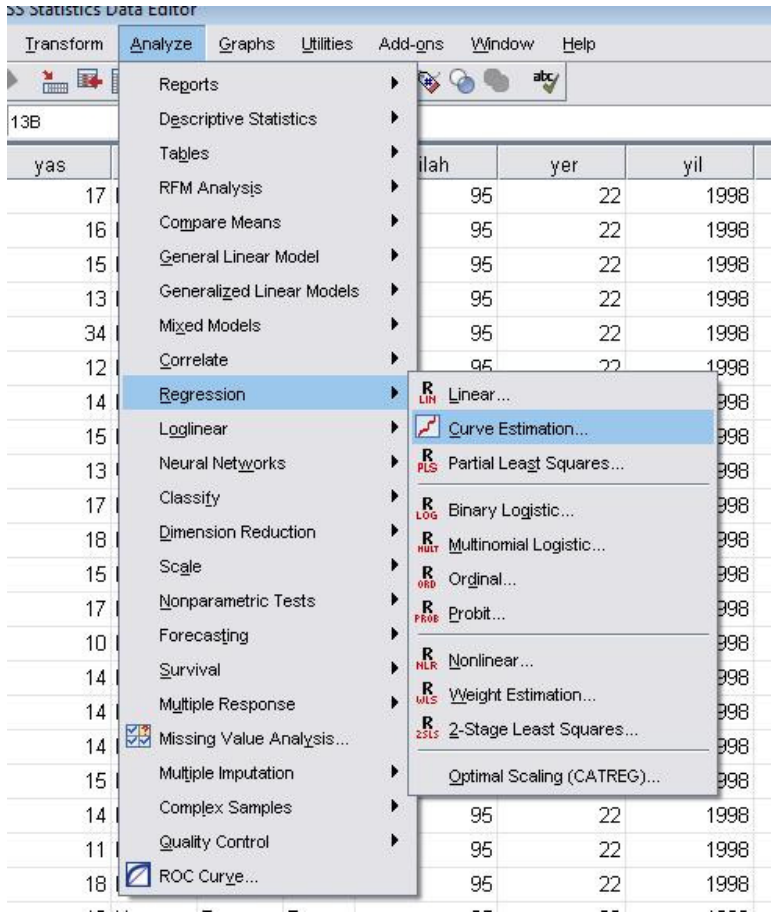


Scatterplot



Görüldüğü gibi çıktı, birçok soruya cevap verebilecek nitelikte detaylı ve kapsamlıdır. Bu çıktıdan değişkenlerin bağımlı olup olmadıkları, bağımlılıklarının derecesi ve yönü, korelasyonu, grafikler, gibi, değişkenler arasındaki bağıntı hakkında çok çeşitli bilgi edinilebilir.

Şimdi SPSS ile eğri uydurma (curve fitting/curve estimation) nasıl yapılır onu inceleyelim. Değişkenlerin kesikli (sürekli) değerleri data dosyamızda mevcuttur. Bunları dikkate alarak sürekli bir eğri tahmin edilebilir. Bunun birçok yolu vardır. İnterpolasyon (interpolation) yoluyla çizilen bu fonksiyonlara spline fonksiyonları ya da interpolasyon fonksiyonları denir. Bu tip fonksiyonları SPSS de hesaplayıp grafiklerini çizebilmektedir. Bunu bir örnek üzerinde görelim. Aşağıdaki komutları sırasıyla verelim.



Yukarıdaki ekranda değişkenlerimizi seçiyoruz. Daha sonra eğri modelimizi seçiyoruz. Burada birinci dereceden lineer bir eğri istiyorsak linear, ikinci dereceden bir eğri uydurulmasını istiyorsak quadratic, üçüncü dereceden bir eğri için cubic gibi istediğimiz modelleri seçiyoruz. Bu komutlar sonrasında aşağıdaki çıktıyı alırız.

Model Description

Model Name		MOD_10
Dependent Variable	1	puan_genel
Equation	1	Linear
	2	Quadratic
Independent Variable		Yaş
Constant		Included
Variable Whose Values Label Observations in Plots		Unspecified
Tolerance for Entering Terms in Equations		,0001

Case Processing Summary

	N
Total Cases	381
Excluded Cases ^a	0
Forecasted Cases	0
Newly Created Cases	0

a. Cases with a missing value in any variable are excluded from the analysis.

Variable Processing Summary

	Variables		
	Dependent	Independent	
	puan_genel	Yaş	
Number of Positive Values	381	381	
Number of Zeros	0	0	
Number of Negative Values	0	0	
Number of Missing Values			
	User-Missing	0	0
	System-Missing	0	0

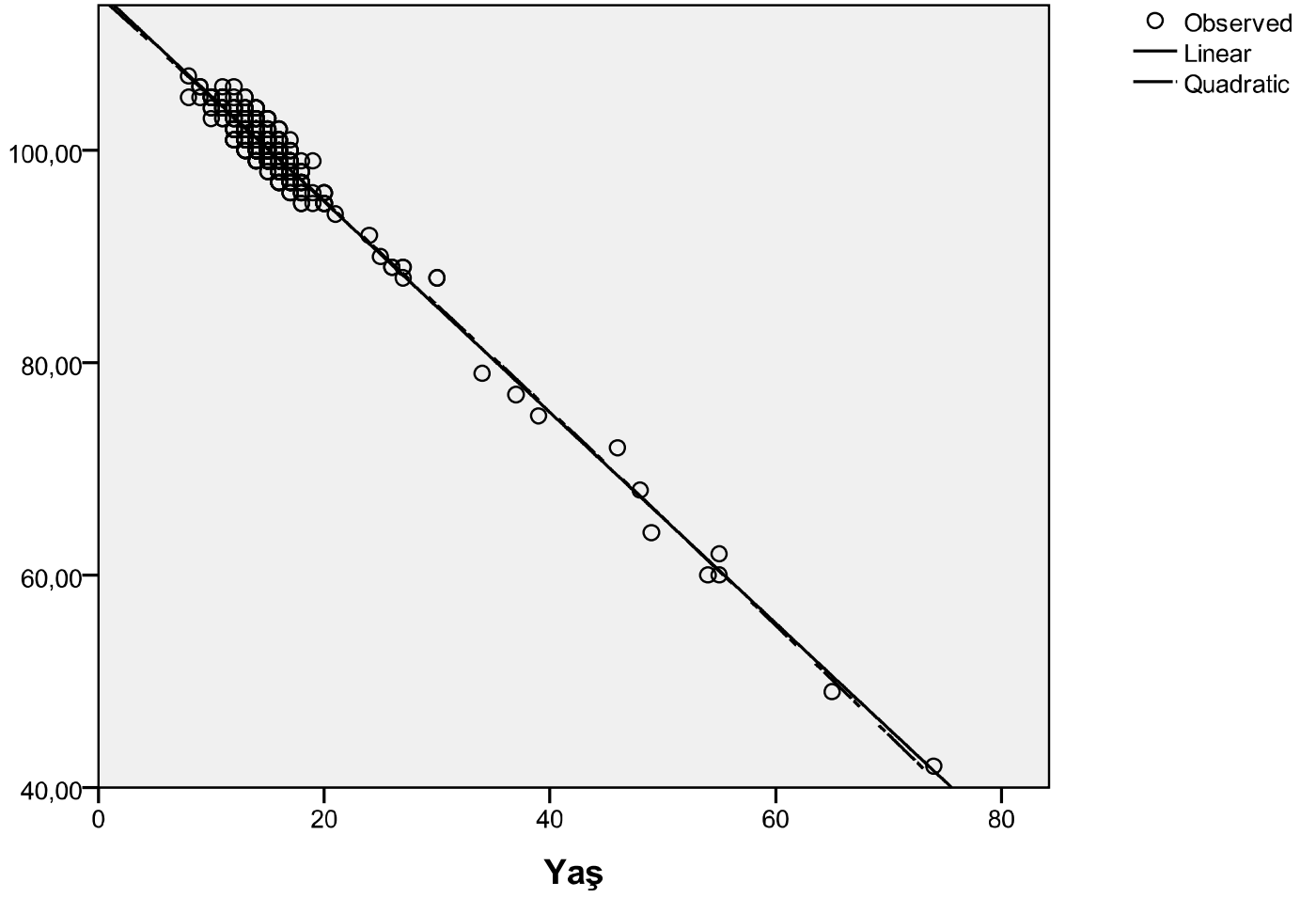
Model Summary and Parameter Estimates

Dependent Variable: puan_genel

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	,966	10622,660	1	379	,000	115,041	-,993	
Quadratic	,966	5313,271	2	378	,000	114,610	-,955	,000

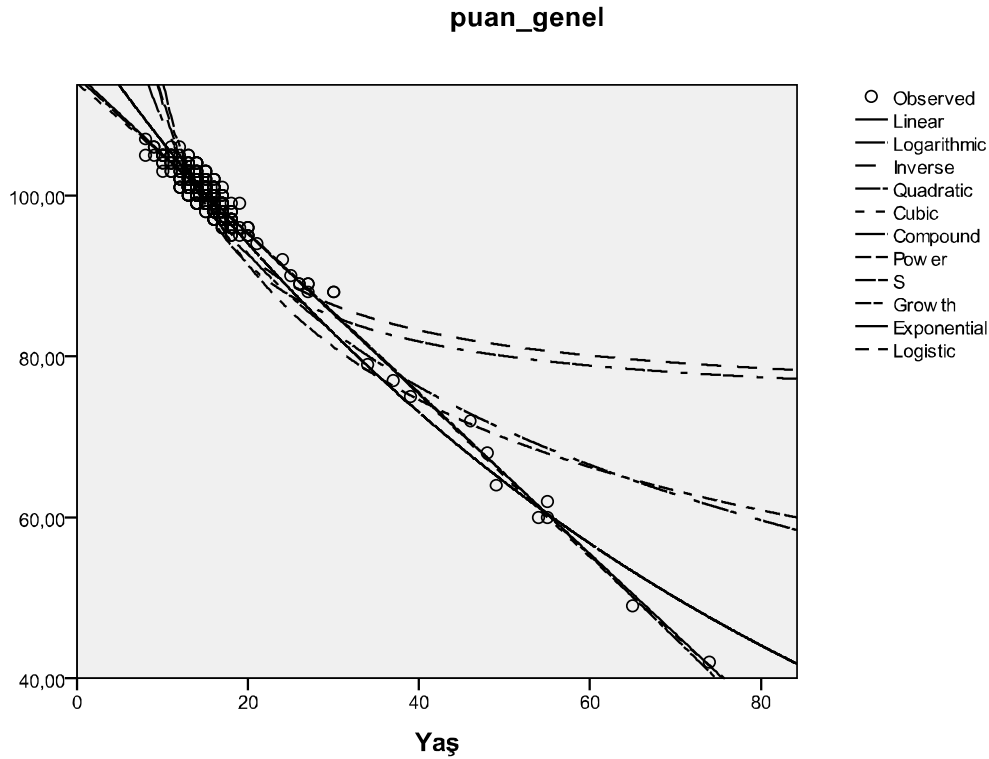
The independent variable is Yaş.

puan_genel

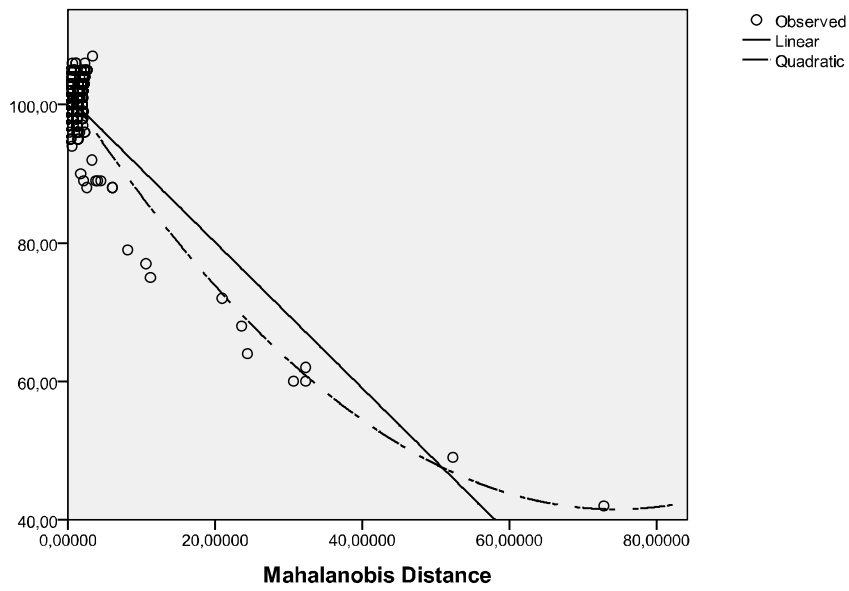


Çıktıdan da görüldüğü gibi, ilk önce değişkenler hakkında bir özet bilgi verilmiş, daha sonra değişkenlerin pozitif, negatif ve sıfır değerlerinin sayıları hakkında bilgi verilmiştir. Daha sonra bu bilgiler dikkate alınarak lineer ve kuadratik model için hesaplamalar yapılmıştır. Significance değerleri verilmiştir. Daha sonrada grafik çizilmiştir. Bu iki değişkeni daha önce de incelemiştik. Bildiğiniz gibi yaş ve puan değişkenleri birbirine aşırı derecede bağlı idi, korelasyon katsayıları 1 e yakın idi. Burada da bunu görmekteyiz. Grafikte içi boş noktalar data dosyamızdaki değerler, sürekli çizgi lineer eğri modeli, kesikli çizgi de kuadratik eğri modelini vermektedir.

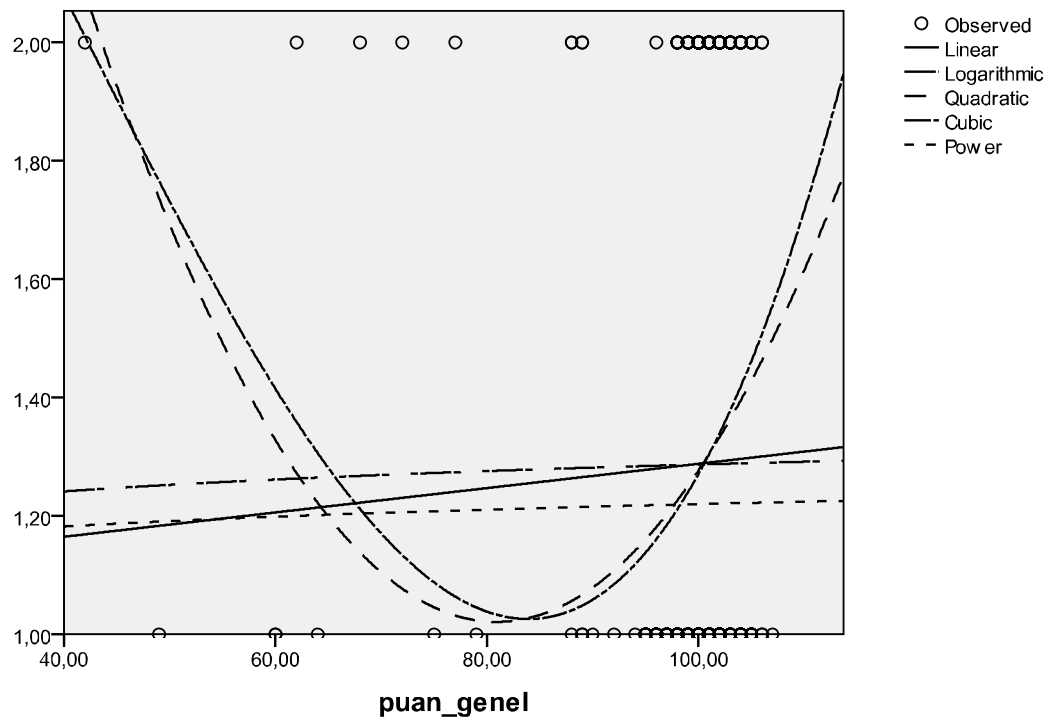
Bu yolla çeşitli değişkenlerin data dosyasındaki bilgileri kullanılarak eğriler uyduralım. Aşağıdaki örnekleri inceleyebiliriz.

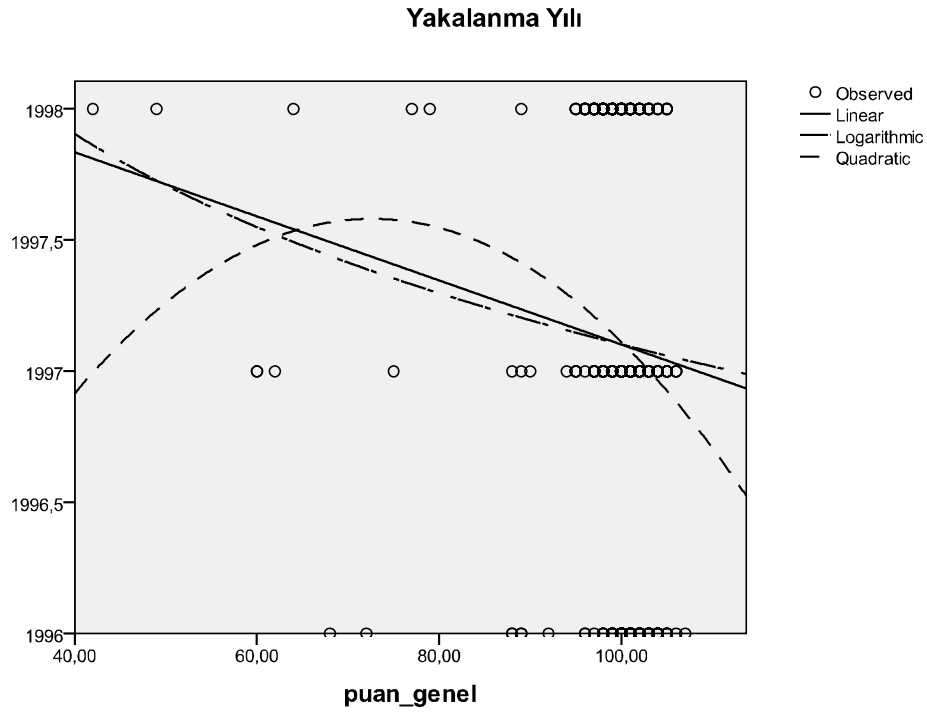
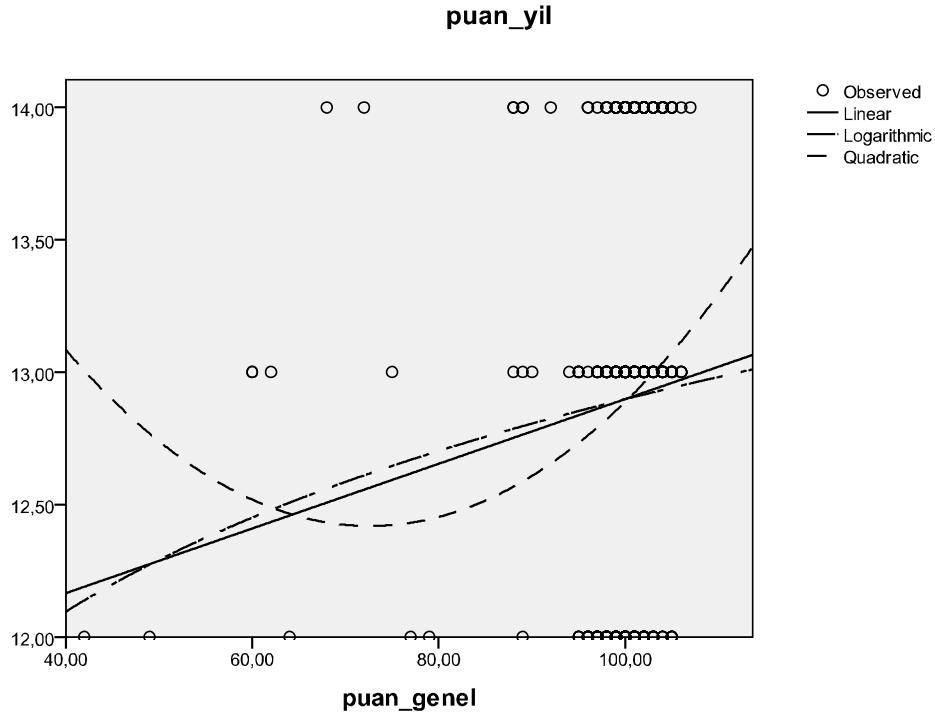


puan_genel



puan_cins





Eğri uydurmada dikkat edilmesi gereken nokta, değişkenlerin dağılımına uygun bir eğri modeli seçmektir.