



İST156 İSTATİSTİĞE GİRİŞ II

HACETTEPE ÜNİVERSİTESİ
İSTATİSTİK BÖLÜMÜ

UYGULAMA 9

Ar. Gör. Dr. Derya Turfan– Ar. Gör. Leyla Bakacak Karabenli

1. 20 bitkinin boy uzunluğu ve yetiştirildiği bölgenin alanı aşağıda verilmiştir.

a. Korelasyon katsayısını hesaplayınız.

b. Doğrusal regresyon denklemini kurup katsayıları yorumlayınız.

c. 100cm²'lik bir çevrede, bitkinin boyunun ne kadar olması beklenir?

Çevre(cm ²)	Boy Uzunluğu(cm)	Çevre(cm ²)	Boy Uzunluğu(cm)
102	50	92	50
99	50	95	46
106	51	109	50
96	49	106	53
100	49	100	50
84	47	99	52
104	48	90	47
97	46	110	52
106	47	116	50
100	52	104	51

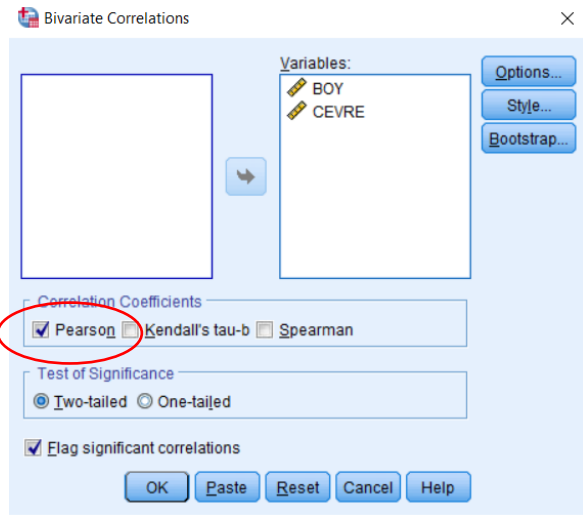
Veri girişi için, SPSS programında *Variable View*'de BOY ve CEVRE olmak üzere 2 farklı sürekli nicel değişken tanımlanmış ve *Data View*'de tabloda verilen değerler girilmiştir.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	BOY	Numeric	8	0		None	None	8	Right	Scale	Input
2	CEVRE	Numeric	8	0		None	None	8	Right	Scale	Input

	BOY	CEVRE
1	50	102
2	50	99
3	51	106
4	49	96
5	49	100
6	47	84
7	48	104
8	46	97
9	47	106
10	52	100
11	50	92
12	46	95
13	50	109
14	53	106
15	50	100
16	52	99
17	47	90
18	52	110
19	50	116
20	51	104

a. Korelasyon katsayısı

Analyze → Correlate → Bivariate



Correlations			
		BOY	CEVRE
BOY	Pearson Correlation	1	,481*
	Sig. (2-tailed)		,032
	N	20	20
CEVRE	Pearson Correlation	,481*	1
	Sig. (2-tailed)	,032	
	N	20	20

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

$H_0: \rho = 0$ (H_0 : Bitkinin boy uzunluğu ile çevresi arasında ilişki yoktur.)

$H_s: \rho \neq 0$ (H_s : Bitkinin boy uzunluğu ile çevresi arasında ilişki vardır.)

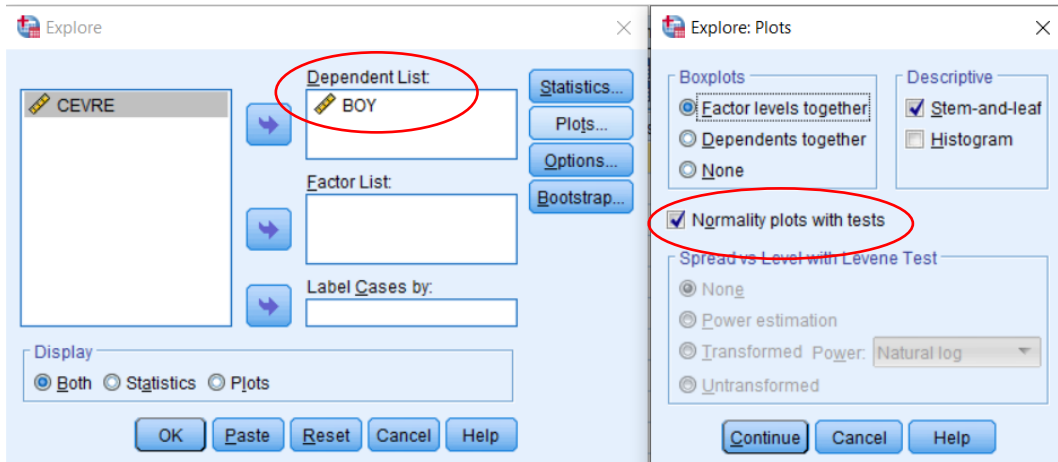
$p=0,032 < \alpha=0,05$ olduğu için H_0 reddedilir ve bitkilerin boy uzunluğu ve çevresi arasında anlamlı bir ilişki olduğu %5 anlamlılık düzeyinde söylenebilir. Boy uzunluğu ve çevre arasında yaklaşık %48'lik pozitif doğrusal bir ilişki vardır. Bitkinin boy uzunluğu artarken çevresi de artmaktadır.

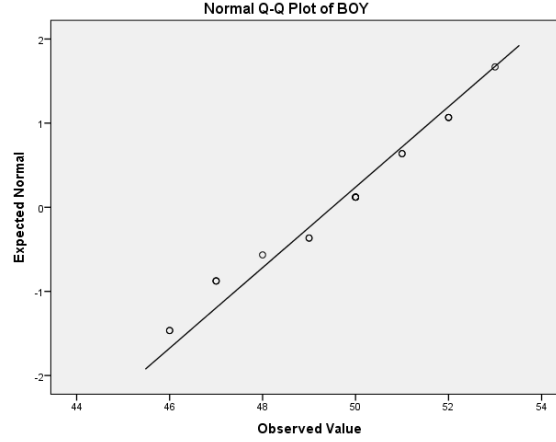
b. Regresyon incelemesi

- Normallik incelemesi

Normallik incelemesi bağımlı değişken olan boy değişkeni üzerinden test edilecektir. Bunun için örneklem hacmi $n=20$, 50'den küçük olduğu için Shapiro-Wilk testi kullanılacaktır. Bu testin öncesinde p-p plot, q-q plot veya box-plot gibi grafiklerden de görsel fayda sağlanabilir.

Analyze → Descriptive Statistics → Explore





Normallik için grafikte noktaların 45 derecelik doğru etrafında yakın bir şekilde dağılması beklenir. Burada da çizgi etrafında çok fazla sapma gözlenmediği için bağımlı değişkenin normal dağılım gösterdiği söylenebilir. Buna destek olarak Shapiro-Wilk testi de incelenecektir.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
BOY	,195	20	,046	,936	20	,197

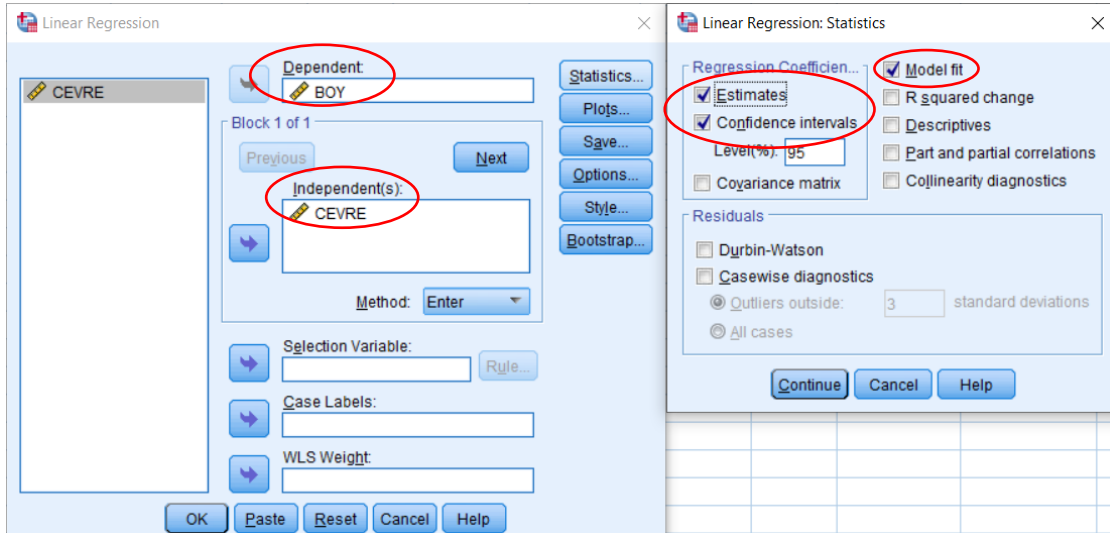
a. Lilliefors Significance Correction

H_0 : Boy uzunluklarının dağılımı ile normal dağılım arasında fark yoktur.

H_s : Boy uzunluklarının dağılımı ile normal dağılım arasında fark vardır.

$p=0,197 > \alpha=0,05$ olduğu için H_0 reddedilemez ve verilerin dağılımının % 5 anlamlılık düzeyinde normal dağılıma uygun dağılış gösterdiği söylenebilir.

Analyze → Regression → Linear



- Belirtme katsayısı

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,481^a	,232	,189	1,882	1,862

a. Predictors: (Constant), CEVRE

b. Dependent Variable: BOY

A şıkında elde edildiği gibi ilişki katsayısı ($R=0,481$) regresyon çıktısında da elde edilmektedir. Belirtme katsayısı ($R^2=0,481^2=0,232$) yani ilişki katsayısının karesi, bağımsız değişkenin bağımlı değişkenin yüzde kaçını açıkladığını ifade etmektedir.

Çevre değişkeni, boy uzunluğunda meydana gelecek değişimin %23'ünü açıklayabilmektedir.

- Varyans çözümleme tablosu ve modelin anlamlılığının testi

ANOVA^a

Model		Sum of Squares (Kareler Toplamı)	df (Serbestlik Derecesi)	Mean Square (Kareler Ortalaması)	F	Sig. (p değeri)
1	Regression (Regresyon)	19,220	1	19,220	5,424	,032^b
	Residual (Artık)	63,780	18	3,543		
	Total	83,000	19			

a. Dependent Variable: BOY

b. Predictors: (Constant), CEVRE

H_0 : Model anlamlı değildir.

H_s : Model anlamlıdır.

$p=0,032 < \alpha=0,05$ veya $F=5,424 > F(0,05; 1; 18)=4,41$ olduğu için H_0 reddedilir ve kurulan regresyon modelinin %95 güvenilrlikle anlamlı olduğu söylenebilir.

- β_0 ve β_1 yorumları ve anlamlılığının testi

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	35,815	5,891		6,080	,000	23,439	48,192
	CEVRE	,136	,058	,481	2,329	,032	,013	,258

a. Dependent Variable: BOY

$\beta_0=35,815 \rightarrow$ Çevrenin etkisi olmadığında bitkinin ortalama boy uzunluğu 35,815 cm'dir.

$\beta_1=0,136 \rightarrow$ Çevrenin etkisi bir birim arttığında bitkinin boy uzunluğunun ortalama 0,136 cm'lik bir artış göstermesi beklenir.

β_0 'ın anlamlılığının testi:

$H_0: \beta_0 = 0$ (Sabit terim anlamlı değildir.)

$H_S: \beta_0 \neq 0$ (Sabit terim anlamlıdır.)

β_0 'ın anlamlı olup olmadığının sonucuna varabilmek için aşağıda verilen üç yöntemden biri tercih edilir.

1. t istatistiği

Hipotezde sıfıra eşit olup olmama durumu test edildiği yani seçenек hipotezi iki yönlü olduğu için t test istatistiği $\alpha/2$ tablo değeri ile karşılaştırılır ($t_{\alpha/2; n-2}$)

$t = 6,080 > t(0,025; 18) = 2,101$ olduğu için H_0 reddedilir.

2. p değeri

$p = 0,000 < \alpha = 0,05$ olduğu için H_0 reddedilir.

3. güven aralığı

Güven aralığı (23,439; 48,192) hipotezde test edilen değeri yani 0'ı içermediği için H_0 reddedilir.

β_1 'ın anlamlılığının testi:

$H_0: \beta_1 = 0$ (Çevrenin boy uzunluğu üzerindeki etkisi anlamlı değildir.)

$H_S: \beta_1 \neq 0$ (Çevrenin boy uzunluğu üzerindeki etkisi anlamlıdır.)

1. t istatistiği

Hipotezde sıfıra eşit olup olmama durumu test edildiği yani seçenек hipotezi iki yönlü olduğu için t test istatistiği $\alpha/2$ tablo değeri ile karşılaştırılır ($t_{\alpha/2; n-2}$)

$t = 2,329 > t(0,025; 18) = 2,101$ olduğu için H_0 reddedilir.

2. p değeri

$p = 0,032 < \alpha = 0,05$ olduğu için H_0 reddedilir.

3. güven aralığı

Güven aralığı (0,013; 0,258) hipotezde test edilen değeri yani 0'ı içermediği için H_0 reddedilir.

Basit doğrusal regresyon modelinde modelin anlamlılığı ile β_1 'in anlamlılığı aynı şeyi ifade etmektedir. Aralarında $t^2=F$ ilişkisi bulunmaktadır ve ANOVA tablosundaki F değeri $2,329^2=5.424$ ile de elde edilebilir. Model anlamlıdır ve çevre değişkeninin boy uzunluğu üzerinde %5 anlamlılık düzeyinde etkisi bulunduğu söylenebilir.

c. Tahmin denklemi: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

X=100 için tahmin değeri $\hat{y} = 35,815 + 0,136 * 100 = 49,415$ olarak elde edilir.

100cm²'lik bir çevrede, bitkinin boyunun ortalama 49 cm olması beklenir.