



SORU 1: 8 öğrencinin istatistik dersinden aldığı puan ile haftada toplam TV izleme süreleri aşağıdaki gibi veriliyor.

Aldığı not	90	80	82	70	60	50	48	33
TV izleme süresi	1	2	4	5	8	7	9	10

- a. Korelasyon katsayısını hesaplayınız.
- b. Doğrusal regresyon denklemini kurup katsayıları yorumlayınız.
- c. Haftada toplam 6 saat TV izlendiğinde, alınacak puanın kaç olması beklenir?

SORU 1-SPSS ÇÖZÜMÜ

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

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	tv_sure	Numeric	8	0		None	None	8	Center	Scale	Input
2	puan	Numeric	8	0		None	None	8	Center	Scale	Input

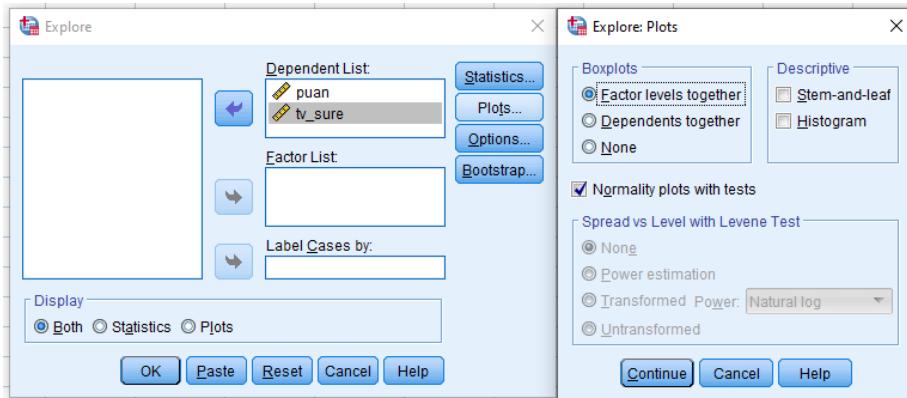
	tv_sure	puan
1	1	90
2	2	80
3	4	82
4	5	70
5	8	60
6	7	50
7	9	48
8	10	33

a. Korelasyon katsayısı

- Normallik incelemesi

Öncelikle her iki değişken için de Normallik incelemesi yapılacaktır. Bunun için örneklem hacmi $n=8$, 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



	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
puan	,165	8	,200*	,958	8	,793
tv_sure	,148	8	,200*	,951	8	,720

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

a. Lilliefors Significance Correction

H_0 : Puanların dağılımı ile normal dağılım arasında fark yoktur.

H_s : Puanların dağılımı ile normal dağılım arasında fark vardır.

$p=0,793 > \alpha=0,05$ olduğu için H_0 reddedilemez ve verilerin dağılıminın % 5 anlamlılık düzeyinde normal dağılıma uygun dağılış gösterdiği söylenebilir.

H_0 : TV izleme sürelerinin dağılımı ile normal dağılım arasında fark yoktur.

H_s : TV izleme sürelerinin dağılımı ile normal dağılım arasında fark vardır.

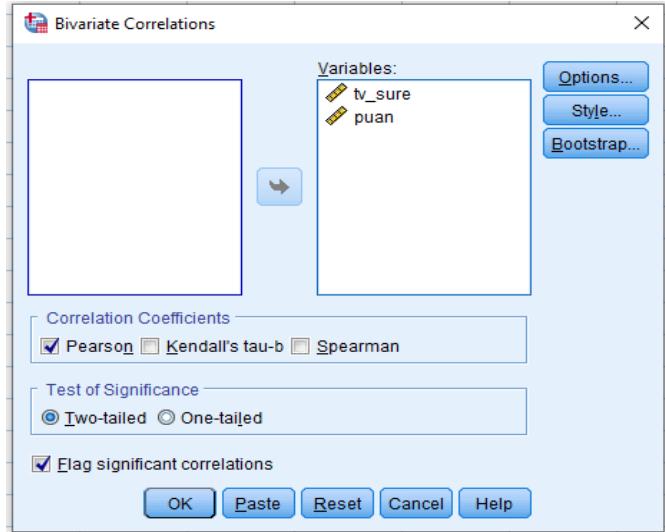
$p=0,720 > \alpha=0,05$ olduğu için H_0 reddedilemez ve verilerin dağılıminın % 5 anlamlılık düzeyinde normal dağılıma uygun dağılış gösterdiği söylenebilir.

Her iki değişken de Normal Dağılım gösterdiği için **Pearson Korelasyon Katsayısı** kullanılacaktır.

$H_0: \rho = 0$ (H_0 : TV izleme süresi ile alınan puan arasında anlamlı bir ilişki yoktur.)

$H_s: \rho \neq 0$ (H_s : TV izleme süresi ile alınan puan arasındaki ilişki anlamlıdır.)

Analyze → Correlate → Bivariate



		Correlations	
		tv_sure	puan
tv_sure	Pearson Correlation	1	-,948**
	Sig. (2-tailed)		,000
	N	8	8
puan	Pearson Correlation	-,948*	1
	Sig. (2-tailed)	,000	
	N	8	8

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

$p=0,000 < \alpha=0,05$ olduğu için H_0 reddedilir ve TV izleme süresi ve alınan puan arasında anlamlı bir ilişki olduğu %5 anlamlılık düzeyinde söylenebilir. TV izleme süresi ve alınan puan arasında yaklaşık %95'lik ters yönlü doğrusal bir ilişki vardır. TV izleme süresi artarken alınan puan azalmaktadır.

b. Regresyon İncelemesi

Burada araştırılan durum alınan puanın TV izleme süresine bağlı olup olmadığıdır.

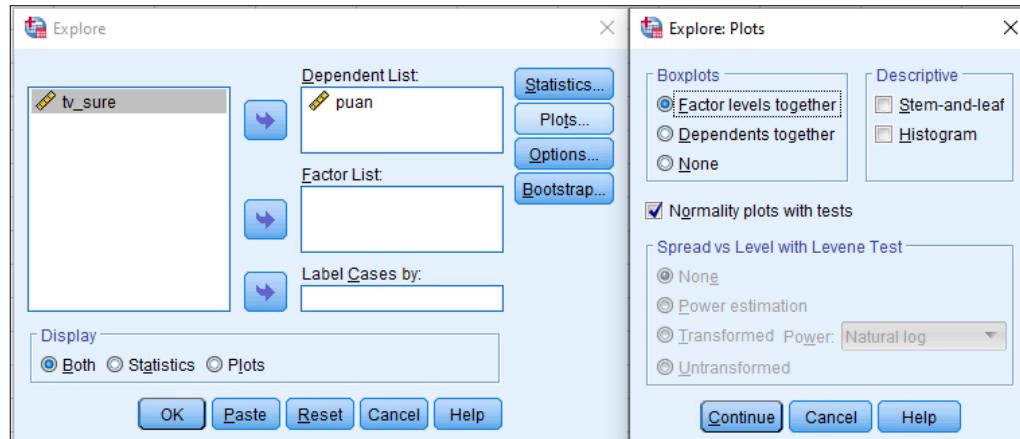
y: puan

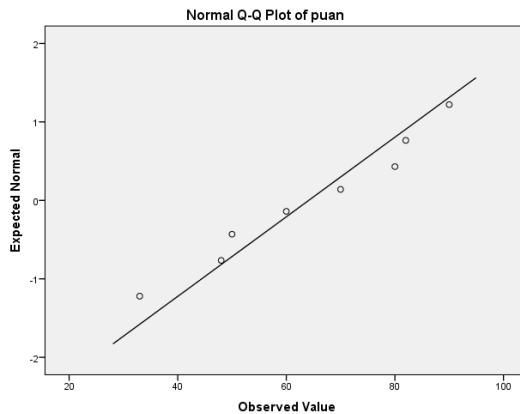
x: tv_sure

- Normallik incelemesi

Normallik incelemesi bağımlı değişken olan **puan** değişkeni üzerinden test edilecektir. Bunun için örneklem hacmi $n=8$, 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 incelenecaktır.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
puan	,165	8	,200*	,958	8	,793

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

a. Lilliefors Significance Correction

H_0 : Puanların dağılımı ile normal dağılım arasında fark yoktur.

H_s : Puanların dağılımı ile normal dağılım arasında fark vardır.

$p=0,793 > \alpha=0,05$ olduğu için H_0 reddedilemez ve verilerin dağılıminin % 5 anlamlılık düzeyinde normal dağılıma uygun dağılış gösterdiği söylenebilir.

Analyze → Regression → Linear

The image shows two overlapping SPSS dialog boxes. The main dialog is "Linear Regression" with the dependent variable "puan" selected and the independent variable "tv_sure" entered. The "Method" is set to "Enter". The "Statistics..." button is highlighted. The second dialog, "Linear Regression: Statistics", contains several options checked under "Estimates", "Confidence intervals", and "Residuals". Under "Residuals", "Outliers outside: 3 standard deviations" is selected. Buttons for "Continue", "Cancel", and "Help" are visible at the bottom of the dialog.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,948 ^a	,899	,882	6,767

a. Predictors: (Constant), tv_sure

Karşımıza çıkan bu tabloda:

R=0,948 → iki değişken arasındaki ilişkinin miktarını vermektedir.

R²=0,948²= 0,899 → Belirtme katsayısı yani ilişki miktarının karesi, bağımsız değişkenin bağımlı değişkenin yüzde kaçını açıkladığını ifade etmektedir.

R² yorumu: TV izleme süresi değişkeni, alınan puanda meydana gelecek değişimin yaklaşık %89'unu açıklayabilmektedir.

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	2446,160	1	2446,160	53,426	,000 ^b
Residual	274,715	6	45,786		
Total	2720,875	7			

a. Dependent Variable: puan

b. Predictors: (Constant), tv_sure

Yukarıdaki tabloda:

Modelin anlamlılığının testine bakılabilir.

H₀: Model anlamlı değildir.

H_S: Model anlamlıdır.

p=0,000 < α=0,05 veya F=53,426 > F(0,05; 1; 6)=5,98 olduğu için H₀ reddedilir ve kurulan regresyon modelinin %95 güvenilirlikle anlamlı olduğu söylenebilir.

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B	
	B	Std. Error				Lower Bound	Upper Bound
1 (Constant)	96,854	5,077		19,078	,000	84,432	109,277
tv_sure	-5,692	,779	-,948	-7,309	,000	-7,598	-3,787

a. Dependent Variable: puan

Yukarıdaki tablodan

$\hat{y} = 96,854 - 5,692x$ modeli yazılabilir ve katsayıların anlamlılığının testine bakılabilir.

$b_0=96,854 \rightarrow$ Bağımsız değişkenin etkisi olmadığında bağımlı değişkenin alacağı değer 96,854'tür.

$b_1 = -5,692 \rightarrow$ Haftalık TV izleme süresi bir birim arttığında alınacak puanın ortalama 5,692 birimlik bir azalma göstermesi beklenir.

β_0 'in 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 olmadığına sonucuna varabilmek için aşağıda verilen üç yöntemden biri tercih edilir.

1. t istatistiği

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

$t = 19,078 > t(0,025; 6) = 2,447$ olduğu için H_0 reddedilir ve model için gereklidir.

2. p değeri

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

3. güven aralığı

Güven aralığı (84,432; 109,277) hipotezde test edilen değeri yani 0'ı içermediği için H_0 reddedilir.

β_1 'in anlamlılığının testi:

$H_0: \beta_1 = 0$ (TV izleme süresinin puan üzerindeki etkisi anlamlı değildir.)

$H_s: \beta_1 \neq 0$ (TV izleme süresinin puan üzerindeki etkisi anlamlıdır.)

1. t istatistiği

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

$t = 7,309 > t(0,025; 6) = 2,447$ olduğu için H_0 reddedilir. β_1 anlamlıdır.

2. p değeri

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

3. güven aralığı

Güven aralığı (-7,598; -3,787) 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 $7,309^2=53,42$ ile de elde edilebilir.

c. Tahmin denklemi: $\hat{y} = b_0 + b_1x$ ya da $\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1x$ olarak yazılabilir.

X=6 için tahmin değeri $\hat{y} = 96,854 - 5,692x = 62,702$ olarak elde edilir.

Haftada toplam 6 saat TV izlendiğinde, alınacak puanın ortalama 63 olması beklenir

SORU 1-R KODLARI

1- Veri girişi

```
aldigi_not <- c(90, 80, 82, 70, 60, 50, 48, 33)  
tv_izleme <- c(1, 2, 4, 5, 8, 7, 9, 10)  
veri <- data.frame(aldigi_not, tv_izleme)
```

2- Normallik testi

```
# normallik için grafik  
qqnorm(veri$aldigi_not)  
# install.packages("ggpubr")  
# library("ggpubr")  
# ggqqplot(aldigi_not)  
  
shapiro.test(veri$aldigi_not)  
# Shapiro-Wilk normality test  
#  
# data: veri$aldigi_not  
# W = 0.95823, p-value = 0.793
```

3- Korelasyon katsayısının hesaplanması

```
r <- cor(veri$aldigi_not, veri$tv_izleme, method = "pearson")  
r  
# -0.9481742
```

4- Korelasyon katsayısının anlamlılığı

```
cor.test(veri$aldigi_not, veri$tv_izleme, method = "pearson", alternative = "two.sided",  
conf.level = 0.95)  
# Pearson's product-moment correlation  
#  
# data: veri$aldigi_not and veri$tv_izleme  
# t = -7.3093, df = 6, p-value = 0.0003346  
# alternative hypothesis: true correlation is not equal to 0  
# 95 percent confidence interval:  
# -0.9908248 -0.7337748  
# sample estimates:  
# cor  
# -0.9481742
```

5- Regresyon incelemesi

```
reg_model <- lm(aldiri_not ~ tv_izleme, data=veri)
summary(reg_model)
# Call:
# lm(formula = aldiri_not ~ tv_izleme, data = veri)
#
# Residuals:
#   Min   1Q   Median   3Q   Max
# -7.0099 -5.8361  0.2219  3.7591  8.6821
#
# Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) 96.8543   5.0768  19.078  1.34e-06 ***
# tv_izleme   -5.6921   0.7787  -7.309  0.000335 ***
# ---
# Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
#
# Residual standard error: 6.767 on 6 degrees of freedom
# Multiple R-squared: 0.899, Adjusted R-squared: 0.8822
# F-statistic: 53.43 on 1 and 6 DF, p-value: 0.0003346
```

5- ANOVA Tablosu

```
anova(reg_model)
# Analysis of Variance Table
#
# Response: aldiri_not
#             Df Sum Sq Mean Sq F value Pr(>F)
# tv_izleme  1 2446.16 2446.16 53.426 0.0003346 ***
# Residuals  6 274.72  45.79
# ---
# Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
```

SORU 2: Bu soruda kullanılacak veri kümesi W.S. Gosset'in 1908 yılında yayınladığı orjinal t-testi makalesinden alınmıştır.

Veri kümesine R üzerinden sleep kodu ile ulaşılabilir.

Çalışma kapsamında iki grup hasta tarafından iki farklı uyku ilacı kullanılmıştır ve "extra" değişkeni ilaçların kullanımı sonrası grplarda uyku saatlerindeki artışı göstermektedir. Her grupta 10 hasta vardır. Toplam veri kümesinde 20 hasta bulunmaktadır.

Grup bilgisi göz önüne alınmaksızın uyku saatlerindeki artışın sıfırdan farklı olup olmadığını %95 güven düzeyinde test ediniz.

SORU 2-SPSS ÇÖZÜMÜ

Veri girişi için, SPSS programında *Variable View*'de **uyku** ismiyle sürekli nicel bir değişken tanımlanmış ve *Data View*'de uyku saatlerindeki artış girilmiştir.

	uyku						
1	,70	6	3,40	11	1,90	16	4,40
2	-1,60	7	3,70	12	,80	17	5,50
3	-,20	8	,80	13	1,10	18	1,60
4	-1,20	9	,00	14	,10	19	4,60
5	-,10	10	2,00	15	-,10	20	3,40

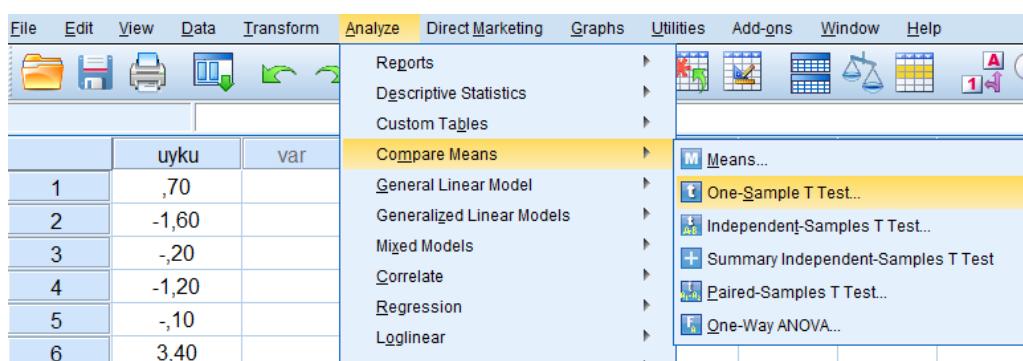
Uyku saatlerindeki artış miktarının sıfırdan farklı olup olmadığını kontrolü için hipotez testi yapılması gerekmektedir. Hipotez testinde de, kitle varyansı bilinmediği ve örneklem hacmi 30'dan küçük olduğu için t testi kullanılacak ve

$$H_0: \mu=0$$

$H_1: \mu \neq 0$ hipotezi test edilecektir.

Kitle varyansı (σ^2) bilinmediği ve $n < 30$ olduğu için t testi kullanılır. Bu testin yapılabilmesi için SPSS'te uygulanması gereken adımlara sırasıyla aşağıda yer verilmiştir.

Analyze → Compare Means → One-Sample T Test



Test değişkeni olarak uyku değişkenin tanımlanması, hipotezde test edilmesi istenen değerin tanımlanması ve anlamlılık değerinin kontrol edilmesi

	uyku	var	var	var	var	var	var
1	,70						
2	-1,60						
3	-,20						
4	-1,20						
5	-,10						
6	3,40						
7	3,70						
8	,80						
9	,00						
10	2,00						
11	1,90						
12	,80						
13	1,10						
14	,10						
15	-,10						
16	4,40						
17	5,50						
18	1,60						

Sonuçların yorumlanması

One-Sample Test						
	Test Value = 0					95% Confidence Interval of the Difference
	t	df	Sig. (2-tailed)	Mean Difference	Lower	
uyku	3,413	19	,003	1,54000	,5956	2,4844

Hipotez testinin sonucuna ulaşabilmek için t test istatistiği, p değeri veya güven aralığı kullanılabilir.

t test istatistiği

Hipotez iki yönlü olduğu için mutlak t test istatistiğinin, $\alpha/2$ tablo değerinden büyük olup olmadığı kontrol edilir.

Eğer $|t| \geq t_{\alpha/2, n-1}$ ise H_0 reddedilir.

$|t|=3,413 > t_{0,025, 19} = 2,093$ olduğu için H_0 reddedilir.

p değeri

$p=0,003 < \alpha=0,05$ olduğu için H_0 reddedilir.

Güven aralığı

Güven aralığının, hipotezde test edilen değeri içerip içermediğini kontrol etmeden önce SPSS çıktısındaki alt ve üst sınıra hipotez test değerinin eklenmesi gerekmektedir.

Güven Aralığı: (0.5956, 2.4844)

Bu işlemden sonra güven aralığı 0 değerini içermediği için H_0 reddedilir.

Yorum: %5 anlamlılık düzeyinde H_0 reddedildiği için, uyku saatlerindeki artış miktarının sıfırdan farklı olduğu söylenebilir (0 olduğuna dair yeterli kanıt bulunmamaktadır).

SORU 2-R KODLARI

1- R üzerinden veri kümesine ulaşılması

```
data <- sleep
```

2- Uyku artışlarına dair kitle ortalaması için hipotez testi

Hipotez testinde kitle ortalamasının 0 (μ) değerine eşit olup olmadığından testi yapılabacağı için seçenek (alternative) hipotezine iki yönlü (two sided) tanımlaması yapılmıştır.

```
t.test(data$extra, mu = 0, alternative = "two.sided", conf.level = 0.95)
```

```
# One Sample t-test
```

```
#
```

```
# data: data$extra
```

```
# t = 3.413, df = 19, p-value = 0.002918
```

```
# alternative hypothesis: true mean is not equal to 0
```

```
# 95 percent confidence interval:
```

```
# 0.5955845 2.4844155
```

```
# sample estimates:
```

```
# mean of x
```

```
# 1.54
```

Yorum için SPSS çıktısı detaylı incelenebilir.