



IST156_1 Dönem Ödevi

ÜNAL GIRAY ERGÜN(2210329023) - BÜNYAMIN AKIN(2210329044)

IST_156_1_ODEV

Bünyamin Akın - Ünal Giray Ergün

2023-05-30

Veri Seti Hakkında Açıklamalar

Bir araştırmaya katılan kişilerin yaşları ve genel olarak hayattan memnuniyet puanları arasındaki ilişkinin araştırıldığı araştırma sonuçları.

Memnuniyet Puanı (Y) : 0 ile 100 arasında bir ölçekten elde edilmiştir.

Yaş (X) : Araştırmaya katılan kişilerin yaşlarıdır.

Verilerin aktarılması.

```
veri <- read.csv('bunyaminakin.csv')  
attach(veri)
```

1) X ve Y değişkeninin her ikisi için ayrı ayrı ortalama ve standart sapmayı hesaplayınız.

1.1- Yaş(X) için ortalama ve standart sapma değerleri:

```
mean(yas)  
## [1] 51.08235  
  
sd(yas)  
## [1] 20.32135
```

1.2- Memnuniyet Puanı (Y) için ortalama ve standart sapma değerleri:

```
mean(puan)  
## [1] 44  
  
sd(puan)  
## [1] 14.81071
```

2) Yaş(X) değişkenini kullanarak kitle ortalaması için %95 güven aralığını hesaplayınız ve yorumlayınız.

1.Yol (MeanCI ile...)

```
# Yaş(X) örneklem büyüklüğü.
length(yas)

## [1] 85

# Gerekli kütüphanelerin aktarılması.
library(DescTools)

## Warning: package 'DescTools' was built under R version 4.2.3

# Kitle ortalaması için %95 güven aralığının bulunması.
MeanCI(yas, conf.level = 0.95)

##      mean   lwr.ci   upr.ci
## 51.08235 46.69914 55.46557
```

2.Yol (İşlem ile...)

```
# Yaş(X) örneklem büyüklüğü.
length(yas)

## [1] 85

varyans <- var(yas)
n <- 85
ort <- mean(yas)
## Tablo degeri hesabi
alpha<-0.05
z.alpha<- qnorm(1-alpha/2)
tablo_degeri_aralik <- c(-z.alpha, z.alpha)
tablo_degeri_aralik

## [1] -1.959964  1.959964

## Guven araligi hesabi
aralik <- c(ort-z.alpha*(sqrt(varyans/n)),ort+z.alpha*(sqrt(varyans/n)))
aralik

## [1] 46.76228 55.40243
```

Yorum: Kitlenin dağılımının normal dağılım olduğu varsayımı altında kitle ortalamasının 46 ile 55 yaş arasında olduğunu %95 güven düzeyinde söyleyebiliriz.

3) Yaş(X) değişkenini kullanarak kitle varyansı için %90 güven aralığını hesaplayınız ve yorumlayınız.

1.Yol (VarCI işlem...)

```
VarCI(yas, conf.level = 0.90)

##      var   lwr.ci   upr.ci
## 412.9574 326.0348 543.0566
```

2.Yol (İşlem ile...)

```
## Tablo değeri
qchisq(c(0.95, 0.05), n-1)

## [1] 106.39484 63.87626

## Güven aralığı
(n-1)*varyans/qchisq(c(0.95, 0.05), n-1)

## [1] 326.0348 543.0566
```

Yorum: Bu güven aralığının kitle varyansını kapsama olasılığı %90'dır. Diğer bir deyişle yaştaki değişimin 326.03 ile 543.05 arasında olduğu %90 güven düzeyinde söylenebilir.

4) Memnuniyet Puanı(Y) değişkenini kullanarak belirlediğiniz bir μ_0 değeri için $H_0:\mu=\mu_0$ hipotezini $H_S:\mu\neq\mu_0$ hipotezine karşı %10 anlamlılık düzeyinde test ediniz ve yorumlayınız.

μ_0 değeri için 50 belirliyorum. $H_0:\mu= 50$ $H_S:\mu\neq 50$ (%10 anlamlılık düzeyinde).

```
# Memnuiyet Puanı(Y) örneklem büyüklüğü.
length(puan)

## [1] 85

# Gerekli kütüphanelerin aktarılması.
library(BSDA)

## Warning: package 'BSDA' was built under R version 4.2.3

## Loading required package: lattice

##
## Attaching package: 'BSDA'

## The following object is masked from 'package:datasets':
##
##      Orange
```

```
# Hipotez Testi
z.test(puan, mu=50, alternative="two.sided", sigma.x=sqrt(var(puan)),
conf.level = 0.90)

##
## One-sample z-Test
##
## data:  puan
## z = -3.735, p-value = 0.0001878
## alternative hypothesis: true mean is not equal to 50
## 90 percent confidence interval:
##  41.35763 46.64237
## sample estimates:
## mean of x
##      44
```

Sonuç:

1. Z test istatistiği: Hipotez iki yönlü olduğu için mutlak Z test istatistiğinin, $\alpha/2$ tablo değerinden büyük olup olmadığı kontrol edilir. Eğer $|Z| \geq Z_{\alpha/2}$ ise H_0 reddedilir. $|Z|=|-3.735|$, $Z_{0.05} = 1.65$, $|Z| \geq Z_{\alpha/2}$ olduğu için H_0 reddedilir.
2. p değeri: p-value = 0.0001878 < $\alpha=0.10$ olduğu için H_0 reddedilir.
3. Güven aralığı: 90 percent confidence interval: 41.35763 46.64237 Güven aralığı 50 değerini içermediği için H_0 reddedilir.

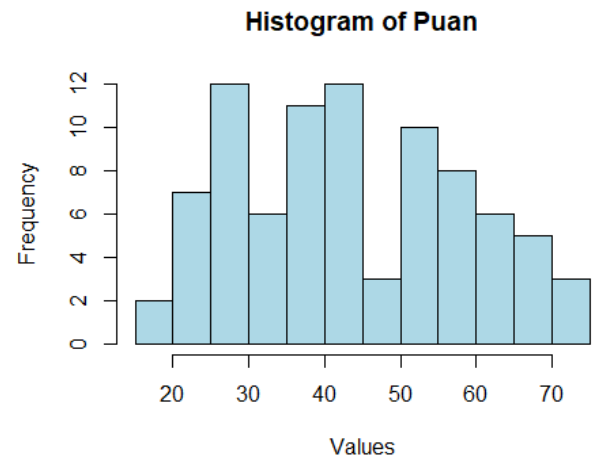
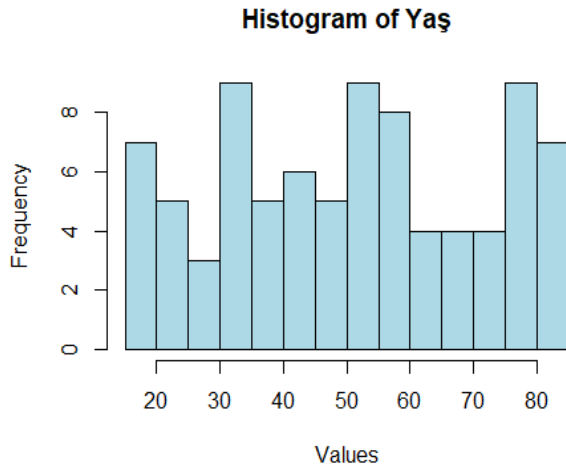
Yorum: %10 anlamlılık düzeyinde H_0 reddedildiği için, kişilerin genel olarak hayattan memnuniyet puanları ortalaması 50 olmadığı söylenebilir. (50 olduğuna dair yeterli kanıt bulunmamaktadır).

5) Yaş(X) ve Puan(Y) değişkenleri arasında Pearson, Spearman ve Kendall'in korelasyon katsayılarından en uygun olanı hesaplayarak yorumlayınız.

5.1- X ve Y değişkenlerinin dağılımına bakalım.

(Verilerin dağılımı)

```
hist(yas, breaks = 10, col = "lightblue", xlab = "Values", ylab =  
"Frequency", main = "Histogram of Yaş")  
  
hist(puan, breaks = 10, col = "lightblue", xlab = "Values", ylab =  
"Frequency", main = "Histogram of Puan")
```



(Normallik Testi)

```
shapiro.test(yas)  
  
##  
##  Shapiro-Wilk normality test  
##  
## data:  yas  
## W = 0.95084, p-value = 0.002688  
  
shapiro.test(puan)  
  
##  
##  Shapiro-Wilk normality test  
##  
## data:  puan  
## W = 0.96053, p-value = 0.01069
```

Yaş için, $p=0,002688 < \alpha=0,05$ olduğu için H_0 reddedilir ve verilerin dağılımının % 5 anlamlılık düzeyinde normal dağılıma uymadığı söylenebilir.

Puan için, $p=0,01069 < \alpha=0,05$ olduğu için H_0 reddedilir ve verilerin dağılımının % 5 anlamlılık düzeyinde normal dağılıma uymadığı söylenebilir.

Değişkenlerimiz normal dağılıma uymuyor bu yüzden Spearman korelasyon katsayısını kullanacağız.

5.2- Korelasyon katsayısını hesaplayalım.

```
r <- cor(puan, yas, method = "spearman")
r
## [1] 0.944743
```

Yaş ve puan arasında pozitif yönlü 0.945'lik güçlü doğrusal bir ilişki vardır.

5.3- Anlamlılığını test edelim.

```
cor.test(puan, yas, alternative = "two.sided", method = "spearman")
## Warning in cor.test.default(puan, yas, alternative = "two.sided", method =
## "spearman"): Cannot compute exact p-value with ties
##
## Spearman's rank correlation rho
##
## data:  puan and yas
## S = 5655, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.944743
```

Yorum: $p < 2.2e-16 < 0.05$, H_0 reddedilir. Yaş ve puan arasında anlamlı ilişki olduğu %95 güven düzeyinde (%5 anlamlılık düzeyinde) söylenebilir ve bu ilişki 0.945'lik pozitif yönlü ilişkidir. Yaş arttıkça puan da artmaktadır.

6) Yaş(X) ve Puan(Y) değişkenleri arasında basit doğrusal regresyon denklemini kurunuz, denklemin anlamlılığını % 5 anlamlılık düzeyinde test ediniz ve yorumlayınız.

6.1- Normallik testi

Daha önceden yaptığımız normallik testine göre verilerimiz normal dağılmıyor fakat ödev için normal dağıldığını varsayıp regresyon denklemini kuracağız.

6.2- Korelasyon ve Belirtme katsayısının hesaplanması

Belirtme katsayısının hesaplayalım daha önceden korelasyon katsayısını hesaplamıştık ($r=0.944743$)

```
R_kare <- r^2
R_kare
```

```
## [1] 0.8925393
```

Yaş değişkeni, puan değişkeninde meydana gelecek değişimin %89'unu açıklayabilmektedir.

6.3- Regresyon incelemesi.

```
reg_model <- lm(puan ~ yas)
summary(reg_model)

##
## Call:
## lm(formula = puan ~ yas)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.7567  -3.8174   0.1219   3.9929  10.9914
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.88573    1.46035   6.085 3.45e-08 ***
## yas         0.68741    0.02658  25.857 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.951 on 83 degrees of freedom
## Multiple R-squared:  0.8896, Adjusted R-squared:  0.8882
## F-statistic: 668.6 on 1 and 83 DF, p-value: < 2.2e-16
```

%5 Anamlık düzeyi için anlamlılık testi

p-value: $< 2.2e-16 < 0.05$ olduğu için H_0 reddedilir.

Model anlamlıdır ve yas değişkeninin puan üzerinde %5 anlamlılık düzeyinde etkisi bulunduğu söylenebilir.

7) 25 ve 40 yaş için yaşamdan memnuniyet puanlarını hesaplayalım.

```
intercept <- 8.88573
slope <- 0.68741

# 25 ve 40 yaş için memnuniyet puanlarını hesaplayalım.
yas_25 <- 25
yas_40 <- 40

puan_25 <- intercept + slope * yas_25
puan_40 <- intercept + slope * yas_40

# Memnuniyet puanlarını yazdıralım.
cat("25 Yaşındaki bireyin memnuniyet puanı:", puan_25, "\n")
## 25 Yaşındaki bireyin memnuniyet puanı: 26.07098
cat("40 Yaşındaki bireyin memnuniyet puanı:", puan_40, "\n")
## 40 Yaşındaki bireyin memnuniyet puanı: 36.38213

Estimate
```

25 Yaşındaki bireyin, memnuniyet puanı 26 olması beklenir.

40 Yaşındaki bireyin, memnuniyet puanı 36 olması beklenir.