

**Açıklama ve uyarılar:** Ödev *el yazısı ile hazırlanacak* ve yanıtlar pdf dosyasına dönüştürüldükten sonra en geç 6 Kasım tarihinde uzak.etu.edu.tr sistemi üzerinden teslim edilecektir. E-posta vb farklı yollarla gönderilen ya da geç teslim edilen ödevler dikkate alınmayacaktır.

## Sorular

1. Regresyon çözümlemesinde kalıntı ölçünlü hatası (KÖH) ve belirleme katsayısı ( $R^2$ ) istatistikleri nasıl yorumlanır? Aralarındaki farkı Açıklayınız.

**Yanıt:** Kalıntı ölçünlü hatası,  $y_i$  tepki değerlerinin regresyondan ortalama kaç birim saptığını verir. Belirleme katsayısı ise tahmin edilen regresyondaki  $x_i$ 'lerin birlikte  $y_i$ 'lerdeki değişikliği ne kadar açıkladığını gösterir. KÖH aslında yakışmanın iyiliğini değil, yakışma eksikliğini anlatır ve birim ölçeğindedir.  $R^2$  ise yakışmayı 0 ve 1 aralığında bir oran olarak ölçer.

2.  $X$  ve  $Y$  arasındaki ilişkinin doğrusal olduğunu ve elimizdeki eğitim verilerine aşağıdaki iki modeli yakıştırdığımızı varsayalım:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$$

İki modelin kalıntı kareleri toplamı (KKT) aynı mıdır, yoksa hangisi daha düşüktür? İki modelden hangisini seçmelisiniz?

**Yanıt:** Daha fazla parametreye sahip olan ikinci model, eğitim veri setinde her zaman daha düşük KKT değeri verir. Ancak hangi modelin seçileceği test verilerindeki performansa bağlıdır.  $X$  ve  $Y$  arasındaki gerçek ilişki eğer doğrusal ise birinci modeli, doğrusal-dışı ise ikinci modeli seçmeliyiz.

3. Elimizdeki  $X_1$ = fiyat (TL),  $X_2$ = reklam harcaması (1000 TL),  $X_3$ = reklam harcaması karesi,  $X_4$ = haftasonu kuklası (haftasonu ise 1) verilerini kullanarak  $Y$ = satış geliri (1000 TL) tepki değişkenini açıklamak istiyoruz. Buna ilişkin regresyon tahminleri aşağıdaki gibidir:

```
lm(formula = satis ~ ., data = Veriler)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.5042	-2.9973	-0.0762	2.8829	10.4517

Coefficients:

Estimate	Std. Error	t value	Pr(> t )
----------	------------	---------	----------

**TOBB - Ekonomi ve Teknoloji Üniversitesi**  
**İKT457 – Ekonomi ve Finans için Yapay Zeka I, İkinci Ödevi**

```
(Intercept) 52.46611    7.07629    7.414 3.20e-10 ***
fiyat      -6.98940    1.05501   -6.625 7.93e-09 ***
reklam     2.91135    0.75901    3.836 0.000286 ***
reklamkare -0.20356    0.03977   -5.118 2.96e-06 ***
haftasonu  10.16296    1.23938    8.200 1.29e-11 ***
---
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.613 on 65 degrees of freedom  
Multiple R-squared: 0.738, Adjusted R-squared: 0.7218  
F-statistic: 45.76 on 4 and 65 DF, p-value: < 2.2e-16

(a) Fiyat değişkeninin katsayısını yorumlayınız.

**Yanıt:** Fiyat değişkeninin katsayısı yaklaşık  $-7$  olarak tahmin edilmiştir. Buna göre ürünün fiyatı 1 TL arttığı zaman diğer her şey sabitken satış gelirlerinin ortalama olarak 7000 TL azalması beklenir.

(b) Reklam değişkeninin katsayısına ait  $t$  istatistiğini ve bunun  $p$ -değerini yorumlayınız.

**Yanıt:** Reklam değişkeninin katsayısına ait  $t$  değeri 3.836'dır. Bunun mutlak değeri olarak 2'den büyük olması tahmin edilen katsayının 0'dan anlamlı derecede uzak olduğunu gösterir.  $P$ -değeri ise  $t$ 'ye ait olasılığın kesin düzeyidir. Buna göre reklam katsayısının 0 olma olasılığı 0,000286'dır. Bulunan oran küçük olduğu için yine,  $H_0 : \beta = 0$  önsavını reddederiz. Sonuç olarak, her iki değer de bize reklam harcamalarının satış üzerinde anlamlı bir etkisi olduğunu söylemektedir.

(c) Belli bir fiyat ve reklam harcaması veriliyken hafta sonları satışların daha fazla olmasını bekler misiniz? Açıklayınız.

**Yanıt:** Evet. Regresyon tahminlerine göre haftasonu değişkeninin katsayısı pozitif ve  $t$  ve  $p$  değerlerine göre de anlamlı bulunmuştur. Sonuçlar, diğer her şey sabitken, hafta sonları satış gelirlerinin ortalama 10 bin TL daha yüksek olacağını söylemektedir.

(d) Regresyona ait  $F$  istatistiğini ve bunun  $p$ -değerini yorumlayınız.

**Yanıt:**  $F$  istatistiği, regresyonun bütün olarak anlamlılığını ölçer. Diğer bir deyişle  $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$  sınamasına ait istatistiktir. Burada bulunan değer 45,76'dır ve bunun  $p$ -değeri de 2,2e-16 olarak hesaplanmıştır.  $P$ -değeri küçük olduğu için regresyonun bütün olarak anlamlı *olmadığı* önsavını *reddederiz*.

4. "Veriler" adlı R veri çerçevesi nesnesine ait değişken adları aşağıda verilmiştir:

```
> names(Veriler)
[1] "satis"      "fiyat"      "reklam"      "reklamkare" "haftasonu"
>
```

(a) Veri setindeki tüm değişkenleri serpilim çizimi matrisi üzerinde gösterecek R komutunu yazınız.

**TOBB - Ekonomi ve Teknoloji Üniversitesi**  
**İKT457 – Ekonomi ve Finans için Yapay Zeka I, İkinci Ödevi**

**Yanıt:**

```
pairs (Veriler)
```

- (b) `lm()` fonksiyonunu kullanarak satışı fiyat ile açıklayan “model1” adındaki basit doğrusal regresyonu tahmin edecek ve `summary()` fonksiyonu ile sonuçlara özetleyecek R komutlarını yazınız.

**Yanıt:**

```
model1 <- lm(satis~fiyat, data=Veriler)
```

- (c) “satis” ve “fiyat” değişkenlerini serpilim çizimi üzerinde gösterecek R komutunu yazınız. `abline()` fonksiyonunu kullanarak regresyon doğrusunu da ekleyiniz.

**Yanıt:**

```
plot (Veriler$fiyat, Veriler$satis)
```

- (d) Fiyat= 5 ve 6 değerleri için model1 kestirimlerini ve bunların güven aralıklarını verecek R komutunu yazınız.

**Yanıt:**

```
predict (model1, data.frame (fiyat=c (5, 6)),  
        interval = "confidence")
```

- (e) `lm()` fonksiyonunu kullanarak satışları veri setindeki diğer tüm değişkenler be bunlara ek olarak fiyat×haftasonu etkileşim değişkeniyle açıklayan “model2” adındaki basit doğrusal regresyonu tahmin edecek R komutunu yazınız.

**Yanıt:**

```
model2 <- lm(satis~.+fiyat:haftasonu, data=Veriler)  
ya da  
model2 <- lm(satis~fiyat+reklam+reklamkare+haftasonu  
            +fiyat:haftasonu, data=Veriler)
```

- (f) Model2’de yeni eklenen değişkenlere ait katsayıların aynı anda 0 olduğunu sıyanan  $F$  testini yapacak R komutunu yazınız.

**Yanıt:**

```
anova (model1, model2)
```