



BÜTÜNLEME SINAV KAĞIDI

Adı:	Dersin Adı: REGRESYON ANALİZİ	Not
Soyadı:	Dersin Kodu: IST3011	
Numarası:	Bölümü: İSTATİSTİK	
İmzası:	Sınav Tarihi: 27/01/2020	

SORULAR

1. (45 puan) Bir tesisde aylık tüketilen elektrik enerjisinin (y (MWh cinsinden)), ortalama ortam sıcaklığı (x_1 (C° cinsinden)) ve çalışılan gün sayısı (x_2) ile ilgili olduğu düşünülmektedir. 12 gözlemden oluşan veri kullanılarak doğrusal regresyon modeli oluşturulmuştur. Bu model için R programından elde edilen sonuçların bir kısmı aşağıda verilmiştir. Bu sonuca göre,

```
lm(formula = y ~ x1 + x2, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-17.051  -9.842   3.167   8.114  13.902

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.9164     81.9126   0.048  0.9629
x1           0.5727     0.2498   2.293  0.0475 *
x2           9.8824     3.7213   2.656  0.0262 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- a) (5 puan) Tahmin edilen doğrusal regresyon modelinin denklemini yazınız.
b) (5 puan) $x_1 = 24C^\circ$ ve $x_2 = 24$ gün iken bir ay için enerji tüketimi ne kadar olur ?
c) (10 puan) Bu model için aşağıda verilen varyans analizi tablosundaki eksiklikleri tamamlayınız.

Değişim Kaynağı	Kareler T.	Serbestlik D.	Kareler Ort.	F-test
Regresyon	5261.2			
Artık				
Toplam	6572.9			

- d) (5 puan) σ^2 nin yansız tahmin edicisini bulunuz.
e) (15 puan) Modeldeki her bir regresyon katsayısının anlamlığını test etmek için gerekli hipotezleri yazınız. $\alpha = 0.05$ olmak üzere regresyon katsayılarının anlamlılığını gerekli hesaplamaları yaparak kısmi t -testi ile test ediniz. Bu model için $X'X$ matrisinin tersi

$$(X'X)^{-1} = \begin{bmatrix} 46.037697 & 0.078063 & -2.071665 \\ 0.078063 & 0.000428 & -0.004212 \\ -2.071665 & -0.004212 & 0.095018 \end{bmatrix}$$

biçiminde elde edilmiştir.

- f) (5 puan) Bu modelde t -testinin kullanabilmesi için hatalar hangi özelliklere sahip olmalıdır ?

2. (55 puan) Bir hastanede 25 hastaya memnuniyet anketi yapılmıştır. Bu veride bağımlı değişken hasta memnuniyeti (y), bağımsız değişkenler hastanın yaşı (x_1), hastalık şiddeti indeksi (x_2), anksiyete indeksi (x_3) ve hastanın tıbbi bir hasta ($x_4 = 0$) ya da cerrahi bir hasta ($x_4 = 1$) olup olmadığını gösteren bir gösterge değişkeni vardır. Bu 4 bağımsız değişkenli model için **R programından elde edilen sonuçların bir kısmı** aşağıda verilmiştir. Bu sonuca göre,

```
Call:
lm(formula = y ~ x1 + x2 + x3 + hasta_tipi, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-18.1322  -3.5662   0.5655   4.7215  12.1448

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 143.8672     6.0437  23.804 3.80e-16 ***
x1           -1.1172     0.1383  -8.075 1.01e-07 ***
x2           -0.5862     0.1356  -4.324 0.000329 ***
x3            1.3064     1.0841   1.205 0.242246
hasta_tipi1   0.4149     3.0078   0.138 0.891672
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.207 on 20 degrees of freedom
Multiple R-squared:  0.9036,    Adjusted R-squared:  0.8843
F-statistic: 46.87 on 4 and 20 DF,  p-value: 6.951e-10
```

a) (13 puan) Tahmin edilen doğrusal regresyon modelinin denklemini hasta tipine göre ayrı ayrı yazınız. Bu denklemlere göre hastanın cerrahi ya da tıbbi hasta olmasının memnuniyeti üzerindeki etkisinin yorumlayınız.

b) (10 puan) Bu model için (4 bağımsız değişkenli) $SST = 10778.2$, $SSE = 1038.9$ olmak üzere varyans analizi tablosunu oluşturunuz.

c) (10 puan) Modelin anlamlılığı için gerekli hipotezleri yazarak, $\alpha = 0.05$ olmak üzere test ediniz.

d) (12 puan) Bu modelden x_4 gösterge değişkenini çıkarırsak elde edilen **yeni model** için $SSE = 1039.9$ dır. Bu yeni model için varyans analizi tablosunu oluşturarak modelin anlamlılığını test ediniz. ($\alpha = 0.05$).

e) (10 puan) $\alpha = 0.05$ için modelden çıkarılan gösterge değişkeninin modele katkısı ölçmek için gerekli hipotezleri yazarak test ediniz.

*Sınav süresi 90 dakikadır. Tüm cevaplarınızı anlaşılır bir biçimde açıklayarak yazınız. Açıklaması olmayan cevaplar değerlendirilmeyecektir. Cep telefonu kullanılması yasaktır.

Formüller: $\sqrt{\hat{\sigma}^2 C_{jj}}$, $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, $\frac{SSE}{n-p}$, $1 - \frac{SSE/(n-p)}{SST/(n-1)}$, $\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t_{n-p}$,

$$\frac{(SSR(\boldsymbol{\beta}) - SSR(\boldsymbol{\beta}_1)) / r}{SSE / (n - p)} \sim F_{r, n-p}, \boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k), \boldsymbol{\beta}_1 = (\beta_0, \beta_1, \dots, \beta_{k-r})$$

BAŞARILAR

Doç. Dr. Fatih KIZILASLAN