

 Fen-Edebiyat Fakültesi	FİNAL SINAV KAĞIDI	
Adı:	Dersin Adı: REGRESYON ANALİZİ	Not
Soyadı:	Dersin Kodu: IST3011	
Numarası:	Bölümü: İSTATİSTİK	
İmzası:	Sınav Tarihi: 15/01/2020	

SORULAR

1. (25 puan) Bir veri seti için çoklu doğrusal regresyon modeli oluşturduğunuzu varsayalım. Bu veri için $X'X$ matrisinin tersi

$$(X'X)^{-1} = \begin{bmatrix} 0.894 & -0.028 & -0.018 \\ 0.028 & 0.002 & 0 \\ -0.018 & 0 & 0.001 \end{bmatrix}$$

biçiminde elde edilmiştir.

- a) (5 puan) Bu regresyon modelinde kaç tane bağımsız değişken vardır ? Bu çoklu doğrusal regresyon modelinin denklemini yazınız.
- b) (5 puan) $n = 15$ ve $SSE = 307$ ise σ^2 nin yansız tahmin edicisini bulunuz.
- c) (5 puan) β_1 regresyon katsayısının standart hatasını bulunuz.
- d) (10 puan) β_1 regresyon katsayısının anlamlılığını test etmek için gerekli hipotezleri yazınız. $\alpha = 0.05$ olmak üzere kısmi t -testi kullanarak β_1 ile ilişkili bağımsız değişkenin modele anlamlı katkısının olabilmesi için $\hat{\beta}_1$ 'in alabileceği olduğu değerlerin aralıklarını bulunuz.

2. (8 puan) k bağımsız değişkenli bir çoklu doğrusal regresyon modelinin denklemini yazınız. Bu model için hipotez testi ve güven aralıkları oluşturabilmemiz için gerekli varsayımlar nelerdir ?

3. (8+4 puan) Otomobil yakıt performansı için y galon başına alınan yol (mil), x_1 motor hacmi (inç küp) ve x_{11} vites türü olmak üzere çoklu doğrusal regresyon modeli oluşturulmuştur. Eğer araç otomatik vites ise $x_{11} = 1$ ve düz vites ise $x_{11} = 0$ olarak tanımlanmıştır. **R programından elde edilen aşağıdaki sonuca göre** (x_{11} =vites olarak alınmıştır.)

Call:

```
lm(formula = y ~ x1 + vites)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.9153	-1.8882	0.1106	1.7706	6.7829

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	33.618408	1.539505	21.837	< 2e-16 ***
x1	-0.045736	0.008682	-5.268	1.2e-05 ***
vites1	-0.498689	2.228198	-0.224	0.824

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

Residual standard error: 3.115 on 29 degrees of freedom

Multiple R-squared: 0.7727, Adjusted R-squared: 0.757

F-statistic: 49.28 on 2 and 29 DF, p-value: 4.696e-10

- a) Tahmin edilen regresyon denklemini ve vites türlerine göre regresyon denklemlerini yazınız. Otomatik vites ile düz vites araçların yakıt performanslarını bu denkleme göre yorumlayınız.

- b) Vites türü yakıt performansını anlamlı bir biçimde etkiler mi, açıklayınız.

4. (55 puan) Bir çoklu doğrusal regresyon modeli y bir kimyasal ürünün viskozitesini, x_1 sıcaklık (F°) ve x_2 tepki süresi (saat) ile ilişkilendirmek için kullanılmıştır. $n = 15$ gözlemden oluşan veri kullanılmıştır.

a) (5 puan) Regresyon katsayılarının EKK tahmin edicileri $\hat{\beta}_0 = 300$, $\hat{\beta}_1 = 0.85$ ve $\hat{\beta}_2 = 10.40$ olarak bulunmuştur. $x_1 = 100 F^\circ$ ve $x_2 = 2$ saat için ortalama viskozitenin tahminini hesaplayınız.

b) (10 puan) $SST = 1230.5$, $SSE = 120.3$ olmak üzere bu model için varyans analizi tablosunu oluşturunuz.

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

d) (8 puan) R^2 belirtme katsayısı ve R_{adj}^2 düzeltilmiş belirtme katsayısını hesaplayarak ve yorumlayınız.

e) (12 puan) Yeni bir bağımsız değişken x_3 karıştırma oranının bu modele eklendiğini varsayalım. Bu yeni model için $SSE = 117.2$ dir. Bu yeni model için varyans analizi tablosunu oluşturarak modelin anlamlılığını test ediniz. ($\alpha = 0.05$) Ayrıca, R^2 ve R_{adj}^2 değerlerini hesaplayınız.

f) (10 puan) Modele eklenen x_3 bağımsız değişkeninin modele katkısını $\alpha = 0.05$ için test ediniz.

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

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

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

BAŞARILAR

Doç. Dr. Fatih KIZILASLAN

Sorular	1	2	3	4
Puan				

4) Modelin $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, bilgilerden ve 15 gözlem yapılmıştır

a) Tahmin edilen regresyon denklemi $\hat{y} = 300 + 0.85x_1 + 10.40x_2$ bilgileri olur

$x_1 = 100 F^\circ$ ve $x_2 = 2$ saat \Rightarrow ortalama viskozite değeri $\hat{y} = 300 + 0.85 \cdot 100 + 10.40(2) = 405.84 //$

b) $SST = 1230.5$, $SSE = 120.3$, $n = 15$ olmak üzere varyans analizi tablosunu oluşturunuz

Değişken Kaynağı	Koşullar Toplamı	Erkinlik Derecesi	Koşullar Ortalaması	F-test
Regresyon	1110.2	2 (k)	555.1	$\frac{555.1}{2} = 277.55$
Artık	120.3	12 (n-p)	10.025	$\frac{10.025}{12} = 0.8354$
Toplam	1230.5	14 (n-1)		

c) Modelin anlamlılığı için $H_0: \beta_1 = \beta_2 = 0$ ve $H_1: \beta_i \neq 0$ ($i=1,2$) en az birinin hipotezleri test etmeliyiz ($\alpha = 0.05$)

$F_h = \frac{555.1}{2} = 277.55$ ve $F_{2,12,0.05} = 3.89$ olmak üzere $F_h > 3.89$ olduğundan $H_0: \beta_1 = \beta_2 = 0$ hipotezi redd edilir. Böylece en az bir $\beta_i \neq 0$ olur. Model anlamlıdır. //

1) a) $(X'X)^{-1}$ matrisi 3×3 olduğundan $p=3 \Rightarrow k=2$ dir. Yani, bu modelde 2 sınımlı değişken vardır.

Bu modelin denklemleri $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ biçimindedir

b) $n=15, SIF=307 \Rightarrow \hat{\sigma}^2 = \frac{SIF}{n-p} = \frac{307}{15-3} = \frac{307}{12} = 25,58$

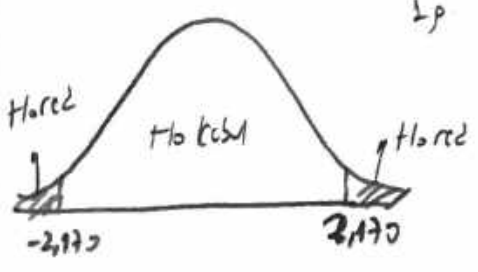
c) $se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 C_{11}}$ ve $C = (X'X)^{-1}$ olduğundan $C_{00}=0,884, C_{11}=0,002, C_{22}=0,001$ dir

Böylece, $se(\hat{\beta}_1) = \sqrt{25,58 \cdot 0,002} = 0,226$

d) $H_0: \beta_1 = 0$ hipotezinin doğru olup olmadığını t -testi ile test ederiz $t = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \sim t_{n-p}$ dir.
 $H_1: \beta_1 \neq 0$ $2p$

Böylece, $\alpha=0,05$ için $H_0: \beta_1 = 0$ hipotezinin reddedilmesi için $t_k = \frac{\hat{\beta}_1 - 0}{0,226}$ $2p$ olmalıdır

$t_{n-p, \alpha/2} = t_{12, 0,025} = 2,179$ olduğundan $t_k > 2,179$ veya $t_k < -2,179$ olmalıdır $2p$



$t_k = \frac{\hat{\beta}_1}{0,226} > 2,179 \Rightarrow \hat{\beta}_1 > 0,492$ olmalıdır $2p$
 veya $t_k < -2,179 \Rightarrow \hat{\beta}_1 < -0,492$ olmalıdır $2p$

2) k sınımlı değişkenli çoklu doğrusal regresyon modeli

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + \epsilon \quad 2p$$

biçimindedir

Çoklu doğrusal regresyon modelinde hipotez testi ve güven aralıkları oluşturabilmek için j'inci verisimleri H_0 hipotezleri $\epsilon_i, i=1, \dots, n$ olmak üzere hataların birbirinden bağımsız ve ortalamalarının $E(\epsilon_i) = 0$ ve varyanslarının sabit $Var(\epsilon_i) = \sigma^2$ olan normal dağılıma sahiptir.

3) R sonuçlarına göre,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_{11} x_{11} = 33,618408 - 0,0457336 x_1 - 0,498689 x_{11} \quad 2p$$

\downarrow motor hızı \downarrow vites hızı
 \downarrow motor hızı \downarrow vites hızı

* Düz vites hızı için ($x_{11} = 0$) tahmin edilen regresyon denklemi:

$$\hat{y} = 33,618408 - 0,0457336 x_1 \quad 2p$$

* Otomatik vites hızı için ($x_{11} = 1$) regresyon denklemi

$$\hat{y} = (33,618408 - 0,498689) - 0,0457336 x_1 \quad 1p$$

* $x_{11} = 0$ dışı diğer değişkenler regresyon denkleminde dışı vitese göre koreleştirebiliriz

Yorum 1
 $y = 33,618 - 0,045736x_1 - 0,408689x_{11}$ denkleme göre

otomatik vitese geçişlerin jütite performansının jübün bşşşına aldığı jöl (mil) dışı vitese göre -0,408689 mil dcha otdır

b) $H_0: \beta_{11} = 0$ hipotezlerin test edilecek ssm R programı jünvündektir tünm f istatistiki ve $H_1: \beta_{11} \neq 0$ $t_h = -0,224$ ve p-değeri = 0,824 kulübün p-değeri $\alpha = 0,05$ olduğundan Ho hipotezi kabul edilir Antemli değileri

4) a) $R^2 = \frac{SSR}{SST} = \frac{1110,2}{1230,5} = \underline{0,9022}$ ve $R^2_{adj} = 1 - \frac{SSE/(n-p)}{SST/(n-1)} = 1 - \frac{(220,3/12)}{(1230,5/14)} = 1 - \frac{10,025}{87,893} = 1 - 0,114 \approx \underline{0,886}$

Yorum 1: Olusturduğumuz regresyon modelr bğımlı değıstkenler toplm değışimn %90'nı (R^2 göre) açıklanmaktadır.

Bunlar olrak $R^2_{adj} = 0,886$ olduğundan bğımlı değıstkenler toplm değışimn %88'i açıklanmaktadır

e) x_3 bğımlı değıstkeni modele eklediğimizde " $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \epsilon$ " bğımlı değıstkeni toplm değışimn %90'nı açıklanmaktadır. olustururuz

Değışim Kaynağı	Kocun Toplamı	Serbestlik Derecesi	Kocun Ort.	F-test
Regresyon	1113,3	3 (k)	372,2	34,84
Artık	117,2	11	20,65	
Toplam	1230,5	14		

$R^2 = \frac{SSR}{SST} = \frac{1113,3}{1230,5} = \underline{0,9047}$
 $R^2_{adj} = 1 - \frac{20,65}{87,893} = 1 - 0,234 \approx \underline{0,766}$
 * (Azalmıştır!!!)

Modelin anlamı: $H_0: \beta_1 = \beta_2 = \beta_3 = 0$
 $H_1: \text{En az bir } \beta_i \neq 0, (i=1,2,3)$

$F_h = 34,84$ ve $F_{3,11,0,05} = 3,59$ olrak $F_h > 3,59$ olduğundan H_0 hipotezi red edilir jün, olusturulan model anlamlıdır

f) x_3 değıstkeni testimn tünm F-testi ile test edilecek $SSR(\beta_0, \beta_1, \beta_2, \beta_3) = 1113,3$ ve $SSR(\beta_0, \beta_1, \beta_2) = 1110,2$ olrak $H_0: \beta_3 = 0$ ve $H_1: \beta_3 \neq 0$ hipotezlern $F_h = \frac{[SSR(\beta_0, \beta_1, \beta_2, \beta_3) - SSR(\beta_0, \beta_1, \beta_2)] / 1}{SSE / (n-p)} = \frac{(1113,3 - 1110,2) / 1}{10,65} = \underline{0,2911}$ bulunur

$F_h = 0,2911$ ve $F_{1,11,0,05} = 4,84$ olrak $F_h < 4,84$ olduğundan $H_0: \beta_3 = 0$ hipotezi kabul edilir. Böylece, x_2 ve x_1 değıstkenleri modele iten x_3 değıstkeninin modele anlamlı bir katkı jütür