 Fen-Edebiyat Fakültesi	ARA SINAV KAĞIDI		
	Adı:	Dersin Adı: REGRESYON ANALİZİ	Not
	Soyadı:	Dersin Kodu: IST3011	
	Numarası:	Bölümü: İSTATİSTİK	
	İmzası:	Sınav Tarihi: 12/11/2019	
SORULAR			

1. (23 puan) 32 farklı otomobilin yakıt performansı ile ilgili olarak galon başına aldığı yol (y) ile motor silindir hacmi (x) (inç küp olarak) değişkenleri kullanılarak oluşturulan basit doğrusal regresyon modeli $\hat{y} = 33.7 - 0.047x$ olarak tahmin edilmiştir. Ayrıca bu model için varyans analizi tablosu aşağıdaki gibi verilmiştir.

Değişim Kaynağı	Kareler T.	Serbestlik D.	Kareler Ort.	F-test hesap
Regresyon	955.72	1	955.72	$F_{hesap} = \frac{955.72}{9.399} = 102.78$
Artık	281.82	$(n-2) = 30$	9.399	
Toplam	1237.54	$(n-1) = 31$		

Bu verilene göre,

- a) (5 puan) Varyans analizi tablosunu tamamlayınız.
- b) (5 puan) Regresyonun anlamlılığını $\alpha = 0.05$ için test ediniz.
- c) (5 puan) Bu modele göre motor hacmi 275 inç küp olan aracın yakıt performansını tahmin ediniz.
- d) (8 puan) Bu modele göre motor hacmi 275 inç küp olan aracın ortalama yakıt performansı için %95 güven aralığı bulunuz. ($\bar{x} = 300, s_{xx} = 250$)
2. a) (9 puan) $y = \beta_0 + \beta_1 x + \epsilon$ biçimindeki basit doğrusal regresyon modelinin varsayımları nelerdir? Ayrıca, regresyon katsayıları için hipotez testi ve güven aralıkları oluşturabilmemiz için gerekli varsayımlar nelerdir?
- b) (6 puan) i) $y = \beta_0 + \beta_1 x + \epsilon$, ii) $y = \beta_0 + \beta_1 x^2 + \epsilon$, iii) $y = \beta_0 + \beta_1 \ln x + \epsilon$ basit regresyon modellerinden hangisi/hangileri doğrusaldır? Açıklayınız.

1) b) $F_{hesap} = 102.78$, $F_{2, n-2, \alpha} = F_{2, 30, 0.05} = 4.27 \Rightarrow F_{hesap} > F_{2, 30, 0.05}$ olduğundan

$H_0: \beta_1 = 0$ hipotezi red edilir. Oluşturulan regresyon modeli anlamlıdır.

c) $x_0 = 275 \Rightarrow \hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ tahmin edilen regresyon denkleminde $x_0 = 275$ ism yakıt performansı için tahmin değeri $\hat{y}_0 = 33.7 - 0.047(275) = \underline{20.775}$ bulunur.

d) $\frac{\hat{y}_0 - E(\hat{y}_0 | x_0)}{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}} \right)} \sim t_{n-2}$ kullanılır. $\hat{\sigma}^2 = \frac{\sum (\hat{y}_i - \hat{y}_i)^2}{n-2} = \frac{JSE}{n-2} = \frac{281.82}{30} = 9.39$ ve $t_{30, 0.025} = 2.042$

$\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}} \right) = 9.39 \left(\frac{1}{31} + \frac{(275 - 300)^2}{250} \right) = 9.39 \left(0.03 + 2.5 \right) = 23.76 \Rightarrow \sqrt{23.76} = 4.87 //$

Böylece, $\hat{y}_0 - t_{n-2, \alpha/2} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}} \right)} \leq E(\hat{y}_0 | x_0) \leq \hat{y}_0 + t_{n-2, \alpha/2} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}} \right)}$ olduğundan

$20.775 - (2.042)(4.87) \leq E(\hat{y}_0 | 275) \leq 20.775 + 2.042(4.87)$

Böylece motor hacmi 275 inç küp olan aracın ortalama yakıt performansı için %95 güven aralığı $[10.81, 30.70]$ olarak bulunur.

3. (42 puan) 1 dekarlık tarım arazilerine atılan gübre miktarı (x) ve bu arazilerden hasat edilen ürün miktarı (y) ile ilgili veri aşağıdaki tabloda verilmiştir.

x_i	8	12	15	20	25	40
y_i	20	30	35	45	50	60

Bu veri için,

- (5 puan) Serpilme diyagramı (saçılım grafiği) oluşturunuz.
- (8 puan) $y = \beta_0 + \beta_1 x + \epsilon$ basit doğrusal regresyon modeli için β_0 ve β_1 parametrelerini tahmin ediniz. Tahmin edilen regresyon denklemini yorumlayınız.
- (8 puan) Regresyonun anlamlılığını $\alpha = 0.05$ için t dağılımını kullanarak test ediniz.
- (8 puan) Varyans analizi tablosunu oluşturunuz ve regresyonun anlamlılığını $\alpha = 0.05$ için test ediniz.
- (5 puan) R^2 belirtme katsayısını hesaplayınız ve yorumlayınız.
- (8 puan) $\alpha = 0.05$ için $H_0: \beta_0 = a$ olmak üzere H_0 hipotezinin kabul edilmesini sağlayacak en pozitif a reel sayısını bulunuz.

4. $y = \beta_0 + \beta_1 x + \epsilon$ basit doğrusal regresyon modelinde β_0 kesim noktasının bilindiğini varsayalım ($\beta_0 = c \in R$ gibi).

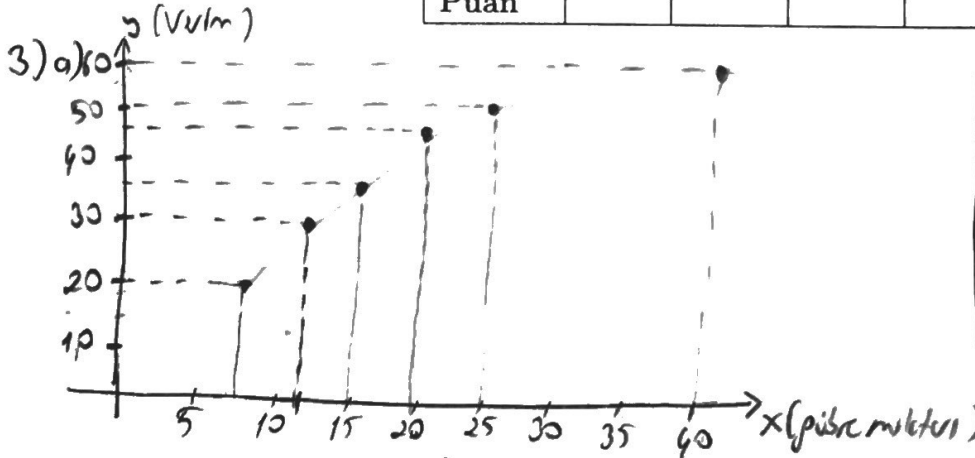
- (10 puan) Bu regresyon modeli için β_1 eğim katsayısının en küçük kareler tahmin edicisini bulunuz.
- (10 puan) Bulduğunuz tahmin edicisi yansız mıdır? Açıklayınız. Varyansını bulunuz.

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

BAŞARILAR

Doç. Dr. Fatih KIZILASLAN

Sorular	1	2	3	4
Puan				



$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2}$$

$$\hat{\beta}_1 = \frac{5555 - 6 \cdot 20 \cdot 40}{3058 - 6 \cdot 20^2}$$

$$\hat{\beta}_1 = \frac{795}{658} = 1,20811$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 40 - 1,208(20)$$

$$\hat{\beta}_0 = 15,84$$

Regresyon denklemi $\hat{y} = 15,84 + 1,208x$

olarak tahmin edilir.

Bu modele göre gübre miktarı 25r. arttığında verimde 1,208br'lik artış meydana gelir.

Gözetim	x_j	y_j	$x_j y_j$	x_j^2	\hat{y}_j	e_j	e_j^2	$(y_j - \bar{y})^2$
1	8	20	160	64	25,504	-5,504	30,294	(-20) ²
2	12	30	360	144	32,336	-0,336	0,113	(-10) ²
3	15	35	525	225	33,560	1,040	1,082	(-5) ²
4	20	45	900	400	40,000	5	25	(5) ²
5	25	50	1250	625	46,040	3,060	15,682	20 ²
6	40	60	2400	1600	64,160	-4,160	17,306	20 ²
Toplam	120	240	5555	3058	240	0	80,477	1050

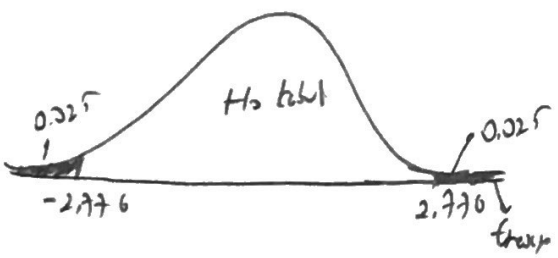
$$\bar{x} = \frac{120}{6} = 20, \bar{y} = \frac{240}{6} = 40$$

c) Regresyonun anlamlılığı ismi \$H_0: \beta_1 = 0\$, \$H_1: \beta_1 \neq 0\$ hipotezler test edilmektedir

$$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}, \hat{\sigma}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} = \frac{SSE}{n-2} = \frac{89.477}{6-2} = 22.369, S_{xx} = 658, \mu(\hat{\beta}_1) = 0.1284$$

$$t = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \sim t_{n-2} \Rightarrow t_{hesap} = \frac{1.208 - 0}{0.1284} = 9.415, t_{n-2, \alpha/2} = t_{4, 0.025} = 2.776$$

\$t_{hesap} = 9.415 > t_{4, 0.025} = 2.776\$ olduğundan \$H_0: \beta_1 = 0\$ hipotezi red edilir.
Olusturulan regresyon modeli anlamlıdır



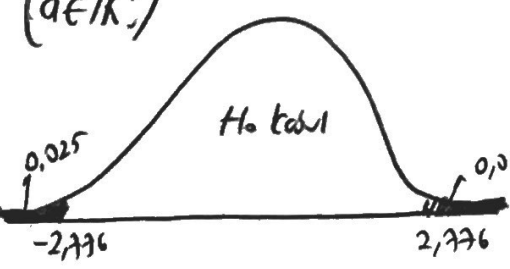
Degisken K.	Kocuk Toplam	Serbestlik Derecesi	Kocuk Ort	\$F_{hesap}\$
Regresyon	560.523	1	560.523	\$\frac{560.523}{22.369} = 42.940\$
Artık	89.477	4	22.369	
Toplam	1050	5		

\$F_{hesap} = 42.940\$
\$F_{2, n-2, \alpha} = F_{2, 4, 0.05} = 7.712\$
\$F_{hesap} > 7.712\$ olduğundan
olusturulan regresyon modeli anlamlıdır

$$SST = \sum_{i=1}^6 (y_i - \bar{y})^2 = 1050, SSE = \sum_{i=1}^6 e_i^2 = 89.477$$

e) \$R^2 = \frac{SSR}{SST} = \frac{560.523}{1050} = 0.534 \Rightarrow\$ (Olusturulan regresyon modelinde)
\$V\$'nin toplam degisiminin %53,4'ü kullanılarak y'ye etkisi ile açıklanmaktadır

f) \$H_0: \beta_0 = a\$, \$H_1: \beta_0 \neq a\$
\$a \in \mathbb{R}^1\$
\$t = \frac{\hat{\beta}_0 - \beta_0}{se(\hat{\beta}_0)} \sim t_{n-2}, se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}\$
\$= \sqrt{22.369 \left(\frac{1}{6} + \frac{20^2}{658} \right)} = \sqrt{27.326} = 5.227\$



$$-2.776 < t_{hesap} = \frac{15.84 - a}{5.227} < 2.776 \Rightarrow -11.554 < 15.84 - a < 11.554$$

\$H_0: \beta_0 = a\$ hipotezi kabul edilir

$$-2.776 < \frac{15.84 - a}{5.227} < 2.776 \Leftrightarrow -11.554 < 15.84 - a < 11.554$$

$$-27.394 < -a < -4.286 \Leftrightarrow 4.286 < a < 27.394$$

Bu aralıkta \$a\$ olursa \$H_0: \beta_0 = a\$ hipotezi kabul edilir.

b) i) Doğrusal degisken
ii) ve iii) modeller doğrusaldır
(\$\beta_0\$ ve \$\beta_1\$ katsayılarının doğrusallık incelenir)

2) a) Doğru doğrusal regresyon modeli (\$y = \beta_0 + \beta_1 x + \epsilon_i, i=1-n\$ olmak üzere) için varsayımlar:
1) \$x\$ bağımsız degisken sabittir. 2) Hatalar \$\epsilon_i, i=1-n\$ için \$E(\epsilon_i) = 0\$ ve \$Var(\epsilon_i) = \sigma^2, i=1-n\$ olan rastgele degiskenlerdir. 3) Hatalar ilistiklidir. 4) Hatalar ile bağımsız degisken ilistiklidir.
Hipotez ve jüri ediktir. Hatalar birbirinden bağımsız ve normal degilime sahip olmalıdır.

3. (42 puan) 1 dekarlık tarım arazilerine atılan gübre miktarı (x) ve bu arazilerden hasat edilen ürün miktarı (y) ile ilgili veri aşağıdaki tabloda verilmiştir.

x_i	8	12	15	20	25	40
y_i	20	30	35	45	50	60

Bu veri için,

- (5 puan) Serpilme diyagramı (saçılım grafiği) oluşturunuz.
- (8 puan) $y = \beta_0 + \beta_1 x + \epsilon$ basit doğrusal regresyon modeli için β_0 ve β_1 parametrelerini tahmin ediniz. Tahmin edilen regresyon denklemini yorumlayınız.
- (8 puan) Regresyonun anlamlılığını $\alpha = 0.05$ için t dağılımını kullanarak test ediniz.
- (8 puan) Varyans analizi tablosunu oluşturunuz ve regresyonun anlamlılığını $\alpha = 0.05$ için test ediniz.
- (5 puan) R^2 belirtme katsayısını hesaplayınız ve yorumlayınız.
- (8 puan) $\alpha = 0.05$ için $H_0 : \beta_0 = a$ olmak üzere H_0 hipotezinin kabul edilmesini sağlayacak en pozitif a reel sayısını bulunuz.

4. $y = \beta_0 + \beta_1 x + \epsilon$ basit doğrusal regresyon modelinde β_0 kesim noktasının bilindiğini varsayalım ($\beta_0 = c \in R$ gibi).

- (10 puan) Bu regresyon modeli için β_1 eğim katsayısının en küçük kareler tahmin edicisini bulunuz.
- (10 puan) Bulduğunuz tahmin edicisi yansız mıdır? Açıklayınız. Varyansını bulunuz.

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

BAŞARILAR

Doç. Dr. Fatih KIZILASLAN

Sorular	1	2	3	4
Puan				

4) $y = \beta_0 + \beta_1 x + \epsilon$ basit doğrusal regresyon modelinde β_0 bilindiğini varsayalım β_1 tahmin edilen regresyon denkleminin $\hat{y} = \beta_0 + \hat{\beta}_1 x$ biçiminde olacaktır. Bu durumda $e_i = y_i - \hat{y}_i = y_i - \beta_0 - \hat{\beta}_1 x_i$, $i=1, \dots, n$ olur (x_i, y_i) , $i=1, \dots, n$ noktaları için $S(\hat{\beta}_1) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i)^2$ hata terimlerinin toplamını minimize eden $\hat{\beta}_1$ EKK tahmin edicisini bulunuz.

$$\frac{\partial S(\hat{\beta}_1)}{\partial \hat{\beta}_1} = 2 \sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i) (-x_i) = 0 \Leftrightarrow \sum_{i=1}^n (y_i x_i - \beta_0 x_i - \hat{\beta}_1 x_i^2) = 0$$

$$\Leftrightarrow \sum_{i=1}^n x_i y_i - \beta_0 \sum_{i=1}^n x_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \Leftrightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \beta_0 \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2}$$

Ayrıca $\frac{\partial^2 S(\hat{\beta}_1)}{\partial \hat{\beta}_1^2} = 2 \sum_{i=1}^n x_i^2 > 0$ her $\hat{\beta}_1$ için olduğundan bu küme için $\hat{\beta}_1$ tahmin edicisi $S(\hat{\beta}_1)$ fonksiyonunun minimum noktasıdır.

b) $E(\epsilon_i) = 0, \text{Var}(\epsilon_i) = \sigma^2, i=1, \dots, n \Rightarrow E(y_i) = \beta_0 + \beta_1 x_i, \text{Var}(y_i) = \sigma^2, i=1, \dots, n$ olur

$$E(\hat{\beta}_1) = \frac{1}{\sum_{i=1}^n x_i^2} \left[\sum_{i=1}^n x_i E(y_i) - \beta_0 \sum_{i=1}^n x_i \right] = \frac{1}{\sum_{i=1}^n x_i^2} \left[\sum_{i=1}^n (x_i \beta_0 + \beta_1 x_i^2) - \beta_0 \sum_{i=1}^n x_i \right] = \beta_1 \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} = \beta_1 //$$

$$\text{Var}(\hat{\beta}_1) = \frac{1}{(\sum_{i=1}^n x_i^2)^2} \sum_{i=1}^n x_i^2 \text{Var}(y_i) = \sigma^2 / \sum_{i=1}^n x_i^2 //$$