BÖLÜM 13

BASİT REGRESYON VE KORELASYON

İstatistiksel araştırmalarda iki veya daha çok değişken arasındaki ilişkinin incelenmesi gerekebilir. Değişkenler arasındaki ilişkinin belirlenmesi için farklı yöntemlerden söz edilebilirse de, en çok kullanılan yöntemler olduklarından ilk olarak akla regresyon ve korelasyon analizi gelmektedir.

Değişkenler arasındaki ilişki matematiksel bir modelle açıklanabileceği gibi, ilişkinin derecesi ve yönü bir katsayı ile de ortaya konabilir. İlişkinin nasıl inceleneceği konusundaki tercih araştırmanın özelliklerine ve araştırmacıya göre değişecektir.

Bu bölümde önce basit doğrusal regresyon, daha sonra doğrusal korelasyon açıklanacak ve bu konuda gerekli açıklamalar yapıldıktan sonra regresyon ve korelasyon arasındaki ilişki ortaya konacaktır.

13.1. BASİT DOĞRUSAL REGRESYON

Değişkenler arasındaki ilişkinin matematiksel bir model ile açıklanmasına regresyon analizi adı verilmektedir. Basit doğrusal regresyondaki basit kelimesi iki değişken arasındaki ilişkinin incelendiğini, doğrusal kelimesi ise değişkenler arasındaki ilişkiyi açıklamak için kullanılacak matematiksel modelin doğrusal olduğunu ifade etmektedir. İki değişken arasındaki ilişkiyi açıklamak için kullanılacak doğrusal modelin matematiksel ifadesi,

$$Y = a + bX$$

şeklinde olacaktır.

Basit doğrusal regresyonda, değişkenler arasındaki ilişki yukarıda verilen fonksiyona benzer bir fonksiyon ile açıklanırken modelin a ve b parametrelerinin belirlenmesi gerekecek ve bunun için değişkenlerin birlikte değişimini gösteren bir bileşik seriden yararlanılacaktır. Değişkenler arasındaki ilişkiyi açıklayan seri çok sayıda değerden oluşacak ve ilişkiyi belirlemek için serinin tümünün analizde kullanılması gerekecektir. Serinin birim sayısını N ile ifade edersek, N birimin oluşturduğu ana kütlenin bir doğrusal regresyon modeli olacaktır. Anakütle regresyon modeli,

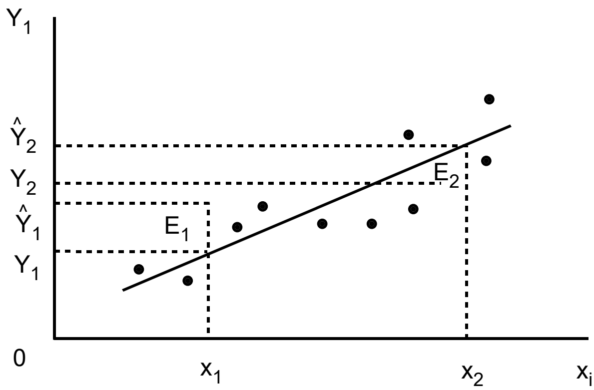
$$Y_i = \alpha + \beta X_i$$

şeklinde yazılırsa bir hata yapılmış olacaktır. Uygulamada karşılaşılabilecek tüm değişkenleri düşünürsek, hangi değişken çifti alınırsa alınsın aralarında tam doğrusal bir ilişki olduğu söylenemez. Değişkenler arasındaki ilişkiyi en iyi açıklayan model doğrusal model olsa bile bazı sapmalar söz konusu olacaktır. Bu nedenle anakütle regresyon modeline hata terimi olarak adlandırılan ve ε_i ile ifade edilen bir değişkenin daha eklenmesi ile modelin,

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

şeklinde ifade edilmesi gerekecektir. Şekil 13.1'de verilen dağılımın noktalarla gösterildiği serpilme diyagramı görülmektedir. Diyagramda yer alan verilerin dağılımına dikkat edilirse bu ilişkinin açıklanması için doğrusal modelin uygun olacağı anlaşılacaktır.

Bu modelin grafiği çizildiğinde elde edilecek doğruya regresyon doğrusu adı verilecektir.



Şekil 13.1

Regresyon doğrusunun bu diyagramda görülen noktalar arasından, değişkenler arasındaki ilişkiyi en iyi açıklayacak şekilde geçirilmesi gerekecektir. Regresyon doğrusu üzerinde yer alacak ve teorik değerler olarak adlandırılan değerler ile gerçek değerler arasındaki fark, hata yani gerçek değerlerden sapmadır. Şekilde gösterilen X_1 değerinin karşılığı olan Y_1 değeri gerçek değer ve regresyon doğrusu üzerinde yer alan $\hat{Y}_1$ değeri ise teorik değerdir. Bu teorik değer gerçek Y_1 değerinden küçük olduğunda ε_1 hata terimi,

$$\varepsilon_1 = Y_1 - \hat{Y}_1 \qquad \text{ve} \qquad Y_1 > \hat{Y}_1$$

olacağından ε_1 pozitif olacaktır. Diğer taraftan X_2 değerinin karşılığı olarak regresyon doğrusu üzerinde yer alan $\hat{Y}_2$ teorik değeri ile gerçek Y_2 değeri için ε_2 hata terimi,

$$\varepsilon_2 = Y_2 - \hat{Y}_2 \qquad \text{ve} \qquad \hat{Y}_2 > Y_2$$

olduğundan ε_2 negatif olacaktır. Görüldüğü gibi hata terimleri pozitif veya negatif işaretli olabilir. Uygulamada verilerin çok sayıda olması; politikaların veya teknolojinin değişmesinin değişkenler arasındaki ilişkiyi etkilemesi gibi nedenlerle anakütle regresyon modelinin parametreleri anakütleden alınacak örnek ile tahmin edilecektir. Bu nedenle N birimli ana kütleden n birimli bir örnek alınacak. Anakütle doğrusal regresyon modelinin örnekten tahminini,

$$Y_i = \hat{\alpha} + \hat{\beta}X_i + e_i$$

olarak gösterilecektir. $\hat{\alpha}$, α 'nın; $\hat{\beta}$, β 'nın tahminini ε_i ise anakütle hata terimi e_i'nin örnekten elde edilecek değerlerini göstermektedir.

Regresyon modelinde Y_i değişkeni etkilenen değişkendir ve bağımlı değişken olarak adlandırılır. Bağımsız değişken olarak adlandırılan X_i değişkeni ise etkileyen, sonuç yaratan değişkendir. Bu nedenle değişkenlerin hangisinin bağımlı hangisinin bağımsız değişken olarak alınacağı önemlidir. Örneğin, gelir, tüketim ilişkisinde gelir tüketimi etkilendiğinden, gelir bağımsız tüketim ise bağımlı değişken olarak ciro, kur ilişkisinde ise ciro karı etkileyeceğinden ciro bağımsız kar bağımlı değişken olarak alınacaktır.

$\hat{\alpha}$, $\hat{\beta}$ daha önce açıklandığı gibi anakütle regresyon modelinin parametrelerinin tahminleridir. $\hat{\alpha}$ 'in tahmin edilen değeri matematiksel olarak $X_i = 0$ olduğundan Y_i'nin alacağı değeri, yani regresyon doğrusunun Y eksenini kesim noktasını gösterecektir. $\hat{\alpha}$ regresyon analizinde bağımsız değişkenin sıfır değerini alması halinde bağımlı değişkenin alacağı değeri göstermektedir. $\hat{\alpha}$ sabit katsayı olarak da adlandırılır.

$\hat{\beta}$ 'in tahmin edilen sayısal değeri matematiksel olarak regresyon doğrusunun eğimini verir $\hat{\beta}$, regresyon analizinde bağımsız değişkenin değerinin 1 birim artması veya azalması karşısında bağımlı değişkenin değerinin ne kadar artacağını veya azalacağını göstermektedir. Bağımsız değişkenin katsayısı olarak da adlandırılan $\hat{\beta}$ değişkenler arasındaki ilişkinin yönünü de gösterir. $\hat{\beta}$ 'in işareti negatif ise değişkenler arasında ters yönlü $\hat{\beta}$'nin işareti pozitif ise değişkenler arasında doğru yönlü, ilişki olacaktır. Değişkenler arasında ilişki yoksa $\hat{\beta}$'in sıfır olması gerekecektir.

Regresyon analizi ile ilgili olarak bu bölümde verilecek formüller ile yapılacak tahminlerin geçerli olması, hata terimleri ile ilgili ve temel varsayımlar olarak adlandırılan varsayımların geçerliliğine bağlıdır. Bu varsayımları şöyle sıralayabiliriz.

- Normallik: Hata terimlerinin dağılımı normal dağılımdır.
- Sıfır Ortalama: Hata terimlerinin dağılımlarının ortalaması sıfırdır.
- Sabit Varyans: Hata terimlerinin her birinin dağılımının varyansı sabittir.
- Otokorelasyon olmaması: Hata terimleri arasındaki ilişkiye otokorelasyon denir. Bu varsayıma göre hata terimleri birbirini etkilemeyecektir.
- X_i değişkeninin tesadüfi değişken olmaması

Bu varsayımlardan bir veya birkaçının geçerli olmaması halinde sonuçlarda sapmalar meydana gelecektir. Varsayımlardan sapmalar olarak adlandırabileceğimiz; varsayımların geçerli olmaması durumu oldukça geniş bir konudur ve burada anlatılmayacaktır.

Örnekten anakütle parametrelerinin tahmininde kullanılacak formülleri elde etmek için istatistiksel tahmin teorisinden yararlanılır ve formüller çeşitli yöntemlerle elde edilebilir. Ancak bu yön-temlerden en çok kullanılanı en küçük kareler yöntemi olduğundan burada parametrelerin tahmini için kullanılacak formüllerin bu yöntem ile elde edilmesine yer verilecektir.

13.1.1. Parametrelerin Tahmini

En küçük kareler yönteminde, parametrelerin tahmininde kullanılacak formüller gerçek ve teorik değerler arasındaki farkların kareleri toplamı minimum olacak şekilde belirlenir. Regresyon analizinde gerçek değerler ile teorik değerler arasındaki farklar hata terimi ile ifade edildiğine göre α ve β hata terimlerinin kareleri toplamı minimum olacak şekilde tahmin edilecektir.

Hata terimlerinin kareleri toplamını T ile ifade edersek,

$$T = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

olacaktır. Teorik değerler

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i \quad Y_i = \alpha + \beta \ X_i$$

denkleminde X_i değerleri yerine konarak elde edileceğinden,

$$T=\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2$$

$$=\sum_{i=1}^{n}\left[Y_i-\left(\alpha+\beta X_i\right)\right]^2$$

$$=\sum_{i=1}^{n}\left[Y_i-\alpha-\beta X_i\right]^2$$

yazılabilir. T'yi minimum yapacak α ve β değerlerini belirlemek için, T'nin α ve β 'ya göre kısmi türevleri alınıp sıfıra eşitlenecektir. Kısmi türevler;

$$\frac{\partial T}{\partial \alpha}=2\sum_{i=1}^{n}\left(Y_i-\alpha-\beta X_i\right)(-1)$$

$$\frac{\partial T}{\partial \beta}=2\sum_{i=1}^{n}\left(Y_i-\alpha-\beta X_i\right)(-X_i)$$

olacaktır.

Kısmi türevleri sıfıra eşitleyip, tahmin oldukları için α ve β 'yı, $\hat{\alpha}$ ve $\hat{\beta}$ ile göstererek gerekli düzenlemeleri yaparsak,

$$2\sum_{i=1}^{n}\left(Y_i-\hat{\alpha}-\hat{\beta} X_i\right)(-1)=0$$

$$2\sum_{i=1}^{n}\left(Y_i-\hat{\alpha}-\hat{\beta} X_i\right)(-X_i)=0$$

$$\sum_{i=1}^{n}Y_i-n\hat{\alpha}-\hat{\beta}\sum_{i=1}^{n}X_i=0$$

$$\sum_{i=1}^{n}X_iY_i-\hat{\alpha}\sum_{i=1}^{n}X_i-\hat{\beta}\sum_{i=1}^{n}X_i^2=0$$

$$\sum_{i=1}^{n}Y_i=n\hat{\alpha}+\hat{\beta}\sum_{i=1}^{n}X_i$$

$$\sum_{i=1}^{n}X_iY_i=\hat{\alpha}\sum_{i=1}^{n}X_i+\hat{\beta}\sum_{i=1}^{n}X_i^2$$

denklemleri elde edilir ve bu denklemler normal denklemler olarak adlandırılmaktadır. α ve β 'nın tahmini için bu iki bilinmeyenli denklem sisteminin çözülmesi gerekmektedir. T'nin ikinci türevleri alınırsa tahmin edilecek değerlerin minimum olacağı görülecektir.

ÖRNEK: Aşağıda verilen serinin basit doğrusal regresyon modelinin parametrelerini tahmin ediniz.

Y_i	X_i	X_iY_i	X_i^2
2	6	12	36
3	7	21	49
5	9	45	81
8	13	104	169
12	15	180	225
30	50	362	560

$$\sum Y_i = n\hat{\alpha} + \hat{\beta}\sum X_i$$

$$\sum X_iY_i = \hat{\alpha}\sum X_i + \hat{\beta}\sum X_i^2$$

$$-10 / 30 = 5\hat{\alpha} + 50\hat{\beta}$$

$$362 = 50\hat{\alpha} + 560\hat{\beta}$$

$$-300 = -50\hat{\alpha} - 500\hat{\beta}$$

$$362 = 50\hat{\alpha} + 560\hat{\beta}$$

$$62 = 0 + 60\hat{\beta}$$

$$\hat{\beta} = \frac{62}{60} = 1{,}03$$

$$30 = 5\hat{\alpha} + 50(1{,}03)$$

$$\hat{\alpha} = \frac{30 - 51{,}5}{5} = -4{,}3$$

İki bilinmeyenli denklemlerin çözümü yerine, birinci denklemden $\hat{\alpha}$ çekilir ve ikinci denklemde yerine konursa,

$$\sum Y_i = n\hat{\alpha} + \hat{\beta}\sum X_i$$

$$\hat{\alpha} = \frac{\sum Y_i - \hat{\beta}\sum X_i}{n} = \overline{Y} - \hat{\beta}\overline{X}$$

$$\sum X_iY_i = \hat{\alpha}\sum X_i + \hat{\beta}\sum X_i^2$$

$$\sum X_iY_i = \left(\overline{Y} - \hat{\beta}\overline{X}\right)\sum X_i + \hat{\beta}\sum X_i^2$$

$$\sum X_iY_i = \overline{Y}\sum X_i + \hat{\beta}\overline{X}\sum X_i + \hat{\beta}\sum X_i^2$$

$$\sum X_iY_i - \overline{Y}\sum X_i = \hat{\beta}(\sum X_i^2 - \overline{X}\sum X_i)$$

$$\hat{\beta} = \frac{\sum X_iY_i - \overline{Y}\sum X_i}{\sum X_i^2 - \overline{X}\sum X_i}$$

pay ve paydayı $\frac{n}{n}$ ile çarparak,

$$\hat{\beta} = \frac{\sum X_iY_i - n\overline{X}\overline{Y}}{\sum X_i^2 - n\left(\overline{X}\right)^2}$$

elde edilir. Bu durum da $\hat{\alpha}$,

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

formülü ile hesaplanabilir. Bu formülleri katsayıların gerçek değerle tahmin formülleri olarak adlandırabiliriz.

ÖRNEK: Normal denklemler ile çözdüğünüz örneği, katsayıların gerçek değerlerden tahmin formülleri ile çözünüz.

$$\sum Y_i = 30 \qquad \sum X_iY_i = 362$$

$$\sum X_i = 50 \qquad \sum X_i^2 = 560$$

olduğundan ortalamalar,

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{30}{5} = 6$$

$$\overline{X} = \frac{\sum X_i}{n} = \frac{50}{5} = 10$$

olacaktır. Değerler formüllerde yerine konursa,

$$\hat{\beta} = \frac{\sum X_iY_i - n\overline{X}\overline{Y}}{\sum X_i^2 - n\left(\overline{X}\right)^2}$$

$$= \frac{362 - 5(10)(6)}{560 - 5(10)^2} = \frac{62}{60} = 1,03 \qquad \hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= 6 - 1,03(10) = -4,3$$

olarak aynı sonuçlar elde edilmiş olur.

Değişkenler için ortalamadan farklar serileri düzenlenerek katsayılar tahmin edilirse işlemler kolaylaşacaktır. Farklar serileri oluşturulduğunda,

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right) = 0$$

$$\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right) = 0$$

olacağını aritmetik ortalamasının özelliklerinden biliyoruz. Normal denklemlerde X_i yerine (X_i - $\overline{X}$) ve Y_i yerine (Y_i - $\overline{Y}$) koyarsak,

$$\sum Y_i = n\hat{\alpha} + \hat{\beta}\sum X_i$$

$$\sum X_i Y_i = \hat{\alpha}\sum X_i + \hat{\beta}\sum X_i^2$$

$$\sum (Y - \overline{Y}_i) = n\hat{\alpha} + \hat{\beta}\sum (X_i - \overline{X})$$

$$\sum (X_i - \overline{X})(Y_i - \overline{Y}) = \hat{\alpha}\sum (X_i - \overline{X}) + \hat{\beta}\sum (X_i - \overline{X})^2$$

Ortalamalardan farkların toplamları sıfır olacağından,

$$\sum (X_i - \overline{X})(Y_i - \overline{Y}) = \hat{\beta}\sum (X_i - \overline{X})^2$$

$$\hat{\beta} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

elde edilir. Bu durumda $\hat{\alpha}$,

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

şeklinde tahmin edilebilir. Bu formüller katsayıların ortalamadan farklar ile tahmin formülleri olarak adlandırılabilirler.

ÖRNEK: Daha önce çözülen örneği ortalamadan farklar formülleri ile çözünüz.

Y_i	X_i	$Y_i - \overline{Y}$	$X_i - \overline{X}$	$(X_i - \overline{X})(Y_i - \overline{Y})$	$(X_i - \overline{X})^2$
2	6	-4	-4	16	16
3	7	-3	-3	9	9
5	9	-1	-1	1	1
8	13	2	3	6	9
12	15	6	5	30	25
30	50	0	0	62	60

$\overline{X} = 10$

$\overline{Y} = 6$

$$\hat{\beta} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

$$= \frac{62}{60} = 1{,}03$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= 6 - 1{,}03\ (10) = -\ 4{,}3$$

olarak aynı sonuçlar bulunur.

Kolaylık olması için $x_i = X_i - \overline{X}$ ve $y_i = Y_i - \overline{Y}$ olarak tanımlarsak,

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

olacaktır.

Gerçek değerlerin toplamları, kareleri ve çarpımlarının toplamları bilindiğinde ortalamadan farklar ile tahmin yapılmak isteniyorsa, ortalamadan farklar serilerinin toplamları kareleri ve çarpımlarının toplamları gerçek değerlerden elde edilebilir.

$$\sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} (X_i^2 - 2\overline{X}X_i + \overline{X}^2)$$

$$= \sum_{i=1}^{n} X_i^2 - 2\overline{X}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} \overline{X}^2$$

$$= \sum_{i=1}^{n} X_i^2 - 2\overline{X}\sum_{i=1}^{n} X_i + n\overline{X}^2$$

$-2\overline{X}\sum_{i=1}^{n} X_i$ ifadesini $\frac{n}{n}$ ile çarparsak,

$$\sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - 2n\overline{X}\frac{\sum X_i}{n} + n\overline{X}^2$$

$$= \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$$

olacaktır. Aynı şekilde,

$$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2$$

olarak elde edilebilir.

$$\sum_{i=1}^{n}(Y_i-\overline{Y})(X_i-\overline{X})=\sum_{i=1}^{n}(X_iY_i-Y_i\overline{X}-\overline{Y}X_i+\overline{XY})$$

$$=\sum_{i=1}^{n}X_iY_i-\overline{X}\sum_{i=1}^{n}Y_i-\overline{Y}\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\overline{XY}$$

$-\overline{X}\sum_{i=1}^{n}Y_i$ ve $-\overline{Y}\sum_{i=1}^{n}X_i$ ifadelerini $\frac{n}{n}$ ile çarparsak,

$$\sum_{i=1}^{n}(Y_i-\overline{Y})(X_i-\overline{X})=\sum_{i=1}^{n}X_iY_i-n\overline{X}\frac{\sum_{i=1}^{n}Y_i}{n}-n\overline{Y}\frac{\sum_{i=1}^{n}X_i}{n}+n\overline{XY})$$

$$=\sum_{i=1}^{n}X_iY_i-n\overline{XY}$$

olarak bulunur.

ÖRNEK: Daha önceki verilen örnekte n =5, $\sum Y_i=$ 30, $\sum X_i$ =50, $\sum X_iY_i$ =362, $\sum X_i^2=$ 560 olarak bilindiğine göre $\sum(X_i-\overline{X})^2$ ve $\sum(X_i-\overline{X})(Y_i-\overline{Y})$ değerlerini hesaplayınız.

$$\overline{X}=\frac{\sum X}{n}=\frac{50}{5}=10$$

$$\overline{Y}=\frac{\sum Y}{n}=\frac{30}{5}=6$$

$$\sum(X_i-\overline{X})^2=\sum X^2-n\overline{X}^2$$

$$= 560-5(10)^2 = 60$$

$$\sum(X_i-\overline{X})(Y_i-\overline{Y})=\sum X_iY_i\ -n\overline{XY}$$

$$= 362 - 5(10)(6) = 62$$

bulunur. Bu sonuçlar daha önce farklar serilerinden elde edilen sonuçlar ile aynıdır.

13.1.2. Teorik Değerlerin Hesaplanması

Bağımsız değişkenin herhangi bir değeri için regresyon modelinden hesaplanacak bağımlı değişkenin değerinin teorik değer olarak adlandırıldığından daha önce söz etmiştik. Diğer bir ifade ile X_i değerinin regresyon modelinde yerine konması ile hesaplanan ve $\hat{Y}_i$ gösterilen değerler teorik değerlerdir. X_i değerleri X_1, X_2,X_3 ,......, X_n ise, teorik değerler,

$$\hat{Y}_1 = \hat{\alpha} + \hat{\beta}(X_1)$$
$$\hat{Y}_2 = \hat{\alpha} + \hat{\beta}(X_2)$$
$$\hat{Y}_3 = \hat{\alpha} + \hat{\beta}(X_3)$$
$$\vdots \quad \vdots \quad \vdots$$
$$\hat{Y}_n = \hat{\alpha} + \hat{\beta}(X_n)$$

olacaktır. Bu değerler regresyon doğrusu üzerinde gerçek Y_i değerlerinin karşılığı olan değerlerdir.

$$e_i = Y_i - \hat{Y}_i$$

olduğu ve regresyonun temel varsayımlarından $\overline{e}$ =0 olacağı bilinmektedir.

$$\overline{e} = \frac{\sum_{i=1}^{n} e_i}{n}$$

olduğundan, ancak

$$\sum_{i=1}^{n} e_i = 0$$

ise $\overline{e} = 0$ olacaktır. Bu nedenle,

$$\frac{\sum_{i=1}^{n} e_i}{n} = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)}{n} = 0$$

olarak yazılabilir. Buradan,

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i) = 0$$

ve

$$\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{Y}_i = 0$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$$

bulunur. Bu sonuçlara göre, gerçek Y_i değerleri ile teorik $\hat{Y}_i$ değerlerinin farklarının toplamı sıfır olacaktır. Ayrıca, gerçek Y_i değerleri toplamı ile teorik $\hat{Y}_i$ değerleri toplamları birbirine eşittir.

ÖRNEK: Daha önce verilen örnek için teorik değerleri hesaplayarak,

$$\sum Y_i = \sum \hat{Y}_i \quad \text{ve} \quad \sum (Y_i - \hat{Y}_i) = 0$$

olduğunu gösteriniz.

Regresyon modeli $Y_i = -4,3 + 1,03X_i$ olarak belirlenmişti.

Y_i	X_i	$\hat{Y}_i$	$e_i = Y_i - \hat{Y}_i$
2	6	1,88	0,12
3	7	2,91	0,09
5	9	4,97	0,03
8	13	9,09	-1,09
12	15	11,15	0,85
30	50	30,00	0

$\hat{Y}_1$ = -4,3 + 1,03 (6) = 1,88

$\hat{Y}_2$ = -4,3 + 1,03 (7) = 2,91

$\hat{Y}_3$ = -4,3 + 1,03 (9) = 4,97

$\hat{Y}_4$ = -4,3 + 1,03 (13) = 9,09

$\hat{Y}_5$ = -4,3 + 1,03 (15) = 11,15

13.1.3. Parametre Tahmincilerinin Varyansları

Regresyon modelinin parametrelerinin varyanslarını hesaplayabilmek için ana kütle hata tahminlerinin varyansının hesaplanması gerekecektir. Anakütle hata terimleri ε_i'lerin varyansı,

$$V(\varepsilon_i) = \sigma^2_{\varepsilon_i} = \frac{\sum_{i=1}^{N} (\varepsilon_i - \overline{\varepsilon})^2}{N}$$

olarak hesaplanacaktır. Regresyonun temel varsayımlarından $\overline{e}$ = 0 olduğunu bilindiğinden

$$\sigma^2_{\varepsilon_i} = \frac{\sum_{i=1}^{N} (\varepsilon_i)^2}{N}$$

olacaktır. Görüldüğü gibi anakütle hata terimlerinin varyanslarını hesaplayabilmek için anakütle hata terimlerinin bilinmesi gerekmektedir. Ana kütle hata terimleri bilinmeyeceğinden, hata terimlerinin varyansının örnekten tahmin edilmesi gerekecektir. Ana kütle hata terimlerinin varyansının örnekten elde edilen tahmini S_e^2 ile gösterilirse, ana kütle hata terimlerinin varyansı örnekten,

$$S_e^2 = \frac{\sum_{i=1}^{n} (\varepsilon_i)^2}{n-2}$$

formülü ile tahmin edilir. Formülde, $\sum e_i^2$ 'nin n yerine (n-2)'ye bölünme sebebi; örnek hata terimleri varyansının, anakütle hata terimleri varyansının sapmalı tahmincisi olmasıdır. Verilen formülle hesaplanan S_e^2 ana kütle hata terimleri varyansının sapmasız tahmincisi olmaktadır.

S_e^2 'nin hesaplanması için hata terimleri kareleri toplamının hesaplanması gerekecektir. Daha önce teorik değerlerin hesaplanmasında açıklandığı gibi, hata terimleri hesaplanarak bunlar yardımı ile hata terimlerinin kareleri toplamı bulunabilir. Hata terimlerinin kareleri toplamının aşağıda verilen formüller ile daha kısa yoldan hesaplanması da mümkündür.

$$Y_i = \hat{\alpha} + \hat{\beta} X_i$$

olduğundan,

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum [Y_i - (\hat{\alpha} + \hat{\beta} X)]^2$$
$$= \sum [Y_i - \hat{\alpha} - \hat{\beta} X_i]^2$$

yazılabilir.

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

olduğundan,

$$\sum e_i^2 = \sum [Y_i - \hat{\alpha} - \hat{\beta} X_i]^2$$
$$= \sum [Y_i - (\overline{Y} - \hat{\beta}\overline{X}) - \hat{\beta} X_i]^2$$

olacaktır. Gerekli düzenlemeleri yaparsak,

$$\sum e_i^2 = \sum [Y_i - \overline{Y} + \hat{\beta}\overline{X} - \hat{\beta} X_i]^2$$
$$= \sum [(Y_i - \overline{Y}) - \hat{\beta}(X_i - \overline{X})]^2$$
$$= \sum [(Y_i - \overline{Y})^2 + \hat{\beta}^2 (X_i - \overline{X})^2 - 2\hat{\beta}(Y_i - \overline{Y})(X_i - \overline{X})]$$
$$= \sum (Y_i - \overline{Y})^2 + \hat{\beta}^2 \sum (X_i - \overline{X})^2 - 2\hat{\beta} \sum (Y_i - \overline{Y})(X_i - \overline{X})$$

olarak bulunur.

$$\hat{\beta} = \frac{\sum (Y_i - \overline{Y})(X_i - \overline{X})}{\sum (X_i - \overline{X})^2}$$

olduğundan,

$$\hat{\beta}\sum (X_i - \overline{X})^2 = \sum (Y_i - \overline{Y})(X_i - \overline{X})$$

olacaktır. Bu eşitlikten yararlanarak,

$$\sum e_i^2 = \sum (Y_i - \overline{Y})^2 + \hat{\beta}\hat{\beta}\sum (X_i - \overline{X})^2 - 2\hat{\beta}\sum (Y_i - \overline{Y})(X_i - \overline{X})$$

$$= \sum (Y_i - \overline{Y})^2 + \hat{\beta}\sum (Y_i - \overline{Y})(X_i - \overline{X}) - 2\hat{\beta}\sum (Y_i - \overline{Y})(X_i - \overline{X})$$

$$= \sum (Y_i - \overline{Y})^2 - \hat{\beta}\sum (Y_i - \overline{Y})(X_i - \overline{X})$$

elde edilir. Ortalamalardan farkları,

$x_i = X_i - \overline{X}$

$y_i = Y_i - \overline{Y}$

ile ifade edersek,

$$\sum e_i^2 = \sum {y_i}^2 - \hat{\beta}\sum x_i y_i$$

olacaktır.

Hata terimlerinin kareleri toplamı ortalamalardan farklar ile hesaplanabildiği gibi, gerçek değerler ile de hesaplanabilir.

$$\sum (Y_i - \overline{Y})^2 = \sum Y^2 - n\overline{Y}^2$$

$$\sum (Y_i - \overline{Y})(X_i - \overline{X}) = \sum X_i Y_i - n\overline{XY}$$

olduğundan bu eşitliklerle,

$$\sum e_i^2 = \sum (Y_i - \overline{Y})^2 - \hat{\beta}\sum (Y_i - \overline{Y})(X_i - \overline{X})$$

$$= \sum Y_i^2 - n\overline{Y}^2 - \hat{\beta}(\sum X_i Y_i - n\overline{XY})$$

olarak yazılabilir.

$$\overline{Y} = \hat{\alpha} + \hat{\beta}\overline{X}$$

olduğundan,

$$\sum e_i^2 = \sum Y_i^2 - n\overline{Y}\overline{Y} - \hat{\beta}\sum X_i Y_i + n\hat{\beta}\overline{XY}$$

$$= \sum Y_i^2 - n\overline{Y}(\hat{\alpha} + \hat{\beta}\overline{X}) - \hat{\beta}\sum X_i Y_i + n\hat{\beta}\overline{XY}$$

$$= \sum Y_i^2 - n\hat{\alpha}\overline{Y} - n\hat{\beta}\overline{X}\overline{Y} - \hat{\beta}\sum X_i Y_i + n\hat{\beta}\overline{X}\overline{Y}$$

$$= \sum Y_i^2 - n\hat{\alpha}\overline{Y} - \hat{\beta}\sum X_i Y_i$$

olarak bulunur.

$$\overline{Y} = \frac{\sum Y_i}{n}$$

olduğundan ifade,

$$\sum e_i^2 = \sum Y_i^2 - n\hat{\alpha}\frac{\sum Y_i}{n} - \hat{\beta}\sum X_i Y_i$$

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$$

şeklini alacaktır.

ÖRNEK: Daha önce verilen örneğin hata terimlerinin varyansını hesaplayınız.

Hata terimleri daha önce hesaplandığından

e_i	e_i^2
0,12	0,0144
0,09	0,0081
0,03	0,0009
-1,09	1,1881
0,85	0,7225
	1,9340

$\sum e_i^2$ =1,9340 olarak bulunur.

Hata terimlerinin kareleri toplamı gerçek değerler ile hesaplandığında,

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$$

$$\sum Y_i^2 = (2)^2 + (3)^2 + (5)^2 + (8)^2 + (12)^2 = 246$$

$$\sum Y_i = 30$$

$$\sum X_i Y_i = 362$$

$$\hat{\alpha} = -4{,}3$$

$$\hat{\beta} = 1{,}03$$

olduğundan,

$$\sum e_i^2 = 246 – (-4{,}3)(30) - 1{,}03(362)$$

$$= 2{,}14$$

olarak bulunur.

Hata terimlerinin kareleri toplamı ortalamalardan farklar ile hesaplandığında,

$$\sum e_i^2 = \sum (Y_i - \overline{Y})^2 - \hat{\beta} \sum (Y_i - \overline{Y})(X_i - \overline{X})$$

Y = 6 olarak bilindiğinden,

$$\sum (Y_i - \overline{Y})^2 = \sum Y^2 - n\overline{Y}^2$$

$$= 246 – 5(6)^2 = 66$$

$$\sum (Y_i - \overline{Y})(X_i - \overline{X}) = 62$$

$$\sum e_i^2 = 66 – 1{,}03(62) = 2{,}14$$

olarak bulunur.

Görüldüğü gibi hata terimlerinin kareleri toplamı hata terimlerinin teker teker hesaplanarak karelerinin alınması halinde 1,934, diğer şekillerde hesaplandığında ise 2,14 olarak bulunmuştur. Aradaki çok küçük ve önemsiz fark, bağımsız değişkenin katsayısının değerinin iki ondalıklı alınmasından kaynaklanmaktadır.

$\sum e_i^2 = 1{,}934$ olduğundan,

$$S_e^2 = \frac{\sum e_i^2}{n-2} = \frac{1{,}934}{5-2} = 0{,}6446$$

$$S_e = \sqrt{0{,}6446} = 0{,}8029$$

olacaktır.

Regresyon modelinin sabit ve bağımsız değişkenin parametrelerinin varyansları,

$$V(\hat{\alpha}) = \sigma_{\hat{\alpha}}^2 = \sigma_{\varepsilon_i}^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right]$$

$$V(\hat{\beta}) = \sigma_{\hat{\beta}}^2 = \frac{\sigma_{\varepsilon_i}^2}{\sum (X_i - \overline{X})^2}$$

formülleri ile hesaplanacaktır. $\sigma_{\varepsilon_i}^2$ bilinmeyip örnekten S_e^2 olarak tahmin edileceğinden, parametrelerin varyanslarını tahmin etmek için,

$$S_{\hat{\alpha}}^2 = S_e^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right]$$

$$S_{\hat{\beta}}^2 = \frac{S_e^2}{\sum (X_i - \overline{X})^2}$$

formülleri kullanılacaktır.

ÖRNEK: Daha önce verilen örnek için parametrelerin varyanslarını hesaplayınız.

$$\overline{X} = 10$$

$$S_e^2 = 0{,}6446$$

$$\sum (X_i - \overline{X})^2 = 60$$

olduğundan,

$$S_{\hat{\alpha}}^2 = S_e^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right]$$

$$= 0{,}6446 \left[\frac{1}{5} + \frac{(10)^2}{60} \right]$$

$$= 1{,}2032$$

$$S_{\hat{\alpha}} = \sqrt{1{,}2032} = 1{,}0969$$

ve

$$S_{\hat{\beta}}^2 = \frac{S_e^2}{\sum (X_i - \overline{X})^2}$$

$$= \frac{0{,}6446}{60}$$

$$= 0{,}0107$$

$$S_{\hat{\beta}} = \sqrt{0{,}0107} = 0{,}1034$$

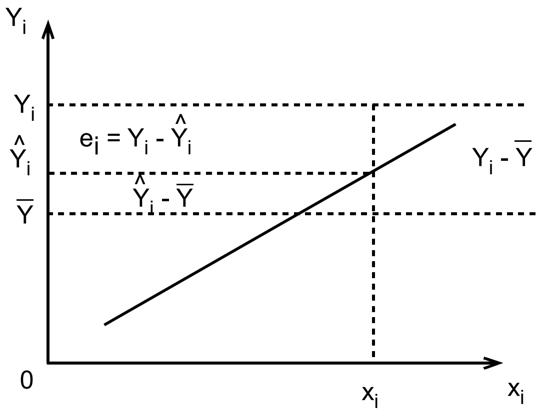
olarak bulunur.

13.1.4. Belirlilik Katsayısı

Bağımlı değişkendeki değişmelerin bağımsız değişken veya bağımsız değişkenler tarafından açıklanma oranını gösteren katsayıya belirlilik katsayısı denmektedir. Basit doğrusal regresyonda tek bağımsız değişken olduğundan, bağımlı değişkendeki değişmeler bu bağımsız değişken tarafından açıklanacaktır. Bağımlı değişkende meydana gelen toplam değişme aritmetik ortalamadan farkların kareleri toplamına eşittir.

$$\text{Toplam Değişme} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} y_i^2$$

Şekil 13.2'de X_i ve Y_i gerçek değerleri ile $\hat{Y}_i$ teorik değeri görülmektedir.



Şekil 13.2

Şekilden de görüldüğü gibi, bağımlı değişkendeki değişme teorik değerin ortalamadan farkı ile hata terimi toplamına eşittir.

$$(Y_i - \overline{Y}) = (\hat{Y}_i - \overline{Y}) + e_i$$

tarafların karelerinin toplamları ise,

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 + \sum e_i^2$$

olarak elde edilir. Burada,

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \text{Toplam Değişmeyi}$$

$$\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 = \text{Açıklanan Değişmeyi}$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \text{Açıklanmayan Değişmeyi}$$

göstermektedir.

Belirlilik katsayısı açıklanan değişmenin toplam değişmeye oranı olarak tanımlandığına göre, R^2 ile ifade edilen belirlilik katsayısı,

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{Y} - \overline{Y})^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}$$

olacaktır. (1- R^2) ise açıklanmayan değişmenin toplam değişmeye oranını verecektir.

Bağımsız değişken bağımlı değişkendeki değişmeleri hiç açıklamıyorsa açıklanan değişme sıfıra eşit, bağımsız değişken bağımlı değişkendeki değişmelerin tümünü açıklıyorsa açıklanan değişme toplam değişmeye eşit olacağından,

$$0 \leq \sum (\hat{Y}_i - \overline{Y})^2 \leq \sum (Y_i - \overline{Y})^2$$

olacaktır. Taraflar toplam değişmeye bölünürse,

$$\frac{0}{\sum (Y_i - \overline{Y})^2} \leq \frac{\sum (\hat{Y}_i - \overline{Y})^2}{\sum (Y_i - \overline{Y})^2} \leq \frac{\sum (Y_i - \overline{Y})^2}{\sum (Y_i - \overline{Y})^2}$$

$$0 \leq R^2 \leq 1$$

olarak bulunur. R^2, 0 ile 1 arasında değer alır. R^2'nin değeri 1'e yaklaştıkça bağımsız değişkenin bağımlı değişkendeki değişmeleri açıklama oranı artar, 0'a yaklaştıkça bağımsız değişkenin bağımlı değişkendeki değişmeleri açıklama oranı azalır.

$$\sum (Y_i - \overline{Y})^2 = \sum (\hat{Y}_i - \overline{Y})^2 + \sum e_i^2$$

olduğundan,

$$\sum (\hat{Y}_i - \overline{Y})^2 = \sum (Y_i - \overline{Y})^2 - \sum e_i^2$$

olacaktır. Bu eşitliği R^2 formülünde yerine koyarsak,

$$R^2 = \frac{\sum (\hat{Y}_i - \overline{Y})^2}{\sum (Y_i - \overline{Y})^2}$$

$$= \frac{\sum (Y_i - \overline{Y})^2 - \sum e_i^2}{\sum (Y_i - \overline{Y})^2}$$

$$= 1 - \frac{\sum e_i^2}{\sum (Y_i - \overline{Y})^2}$$

olarak bulunur. Yani, belirlilik katsayısı açıklanmayan değişmenin toplam değişmeye oranının 1'den çıkarılmasına eşittir. Aynı bağımlı değişken için farklı bağımsız değişkenler ile oluşturulacak regresyon modellerinden belirlilik katsayısı daha büyük olan model ilişkiyi daha iyi açıklayacaktır.

ÖRNEK: Daha önce verilen örneğin belirlilik katsayısını hesaplayarak, bulduğunuz sonucu yorumlayınız.

Daha önce çözdüğümüz örnekte;

$$\sum e_i^2 = 1{,}934$$

ve

$$\sum (Y_i - \overline{Y})^2 = 66$$

olarak bulunmuştu.

Toplam değişme:

$$\sum (Y_i - \overline{Y})^2 = 66$$

Açıklanmayan değişme:

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = 1{,}934$$

olduğundan,

Açıklanan değişme:

$$\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = 66 - 1{,}934 = 64{,}066$$

olacaktır.

Açıklanan değişme yukarıda verilen formül ile hesaplanabildiği gibi seriden de hesaplanabilir.

$\overline{Y}$ = 6 olarak bilinmektedir.

$\hat{Y}_i$	$\hat{Y}_i - \overline{Y}$	$(\hat{Y}_i - \overline{Y})^2$
1,88	-4,12	16,9744
2,91	-3,09	9,5481
4,97	-1,03	1,0609
9,09	3,09	9,5481
11,15	5,15	26,5225
	0	63,654

İki hesaplama şekli arasındaki fark katsayıların değerlerinin tam alınmamasından kaynaklanmaktadır.

$$R^2 = \frac{64{,}066}{66} = 0{,}97$$

veya

$$R^2 = 1 - \frac{1{,}934}{66} = 0{,}97$$

olarak bulunur. Bu sonuca göre, bağımsız değişken bağımlı değişkendeki değişmelerin % 97'sini açıklamaktadır. 1- 0,97 = 0,03 olduğundan bağımsız değişken bağımlı değişkendeki değişmelerin % 3'ünü açıklayamamakta ve bu değişme diğer değişkenler tarafından açıklanmaktadır.

13.1.5. Katsayıların Testi

Regresyon modelinin tahmin edilen katsayılarının istatistiksel olarak anlamlı olup olmadığı test edilebilir. Bu amaçla t ve F testi olmak üzere iki farklı test uygulanabilir. F testinin basit regresyonda uygulanması anlamlı olmasa da burada kısaca açıklanacaktır.

13.1.5.1. t -Testi

Katsayıların ayrı ayrı istatistiksel anlamlılığı t testi ile incelenir. Basit regresyonda sabit ve eğim katsayıları olduğuna göre bunlara ayrı ayrı test uygulanır. Katsayıların anlamlılığı sözkonusu olduğundan, sıfırdan farklı olup olmadıklarına belirlenen hata payı ile karar verilecektir.

Katsayılar test edilirken hipotez testleri bölümünde açıklandığı gibi önce hipotezler oluşturulur. Sabit katsayı test edilirken temel hipotez,

$$H_0 : \alpha = 0$$

karşıt hipotez, test çift taraflı için,

$$H_1 : \alpha \neq 0$$

tek taraflı test için ise,

$$H_1 : \alpha > 0$$

veya

$$H_1 : \alpha < 0$$

şeklinde oluşturulacaktır.

Hipotezler oluşturulduktan sonra karar aşamasında kullanılacak kritik değer veya tablo değeri belirlenir. Kritik değerin belirlenmesinde testin tek veya çift taraflı olması yanında örnek birim sayısı da önemlidir. Örnek birim sayısı 30'dan büyük ise örnekleme dağılımının normal dağılım olduğu kabul edilir ve çift taraflı test için $Z_{\alpha/2}$, tek taraflı test için Z_α değeri normal dağılım tablosundan belirlenen α hata payı ile bulunur.

Örnek birim sayısı 30'dan küçük ise örnekleme dağılımı t dağılımıdır. Basit regresyonda t dağılımının serbestlik derecesi (n-2)'dir. Belirlenen α hata payı ile t dağılımı tablosundan çift taraflı test için $t_{\alpha/2}$, tek taraflı test için t_α değeri bulunacaktır.

Hipotez testlerinde üçüncü adım test istatistiğinin hesaplanmasıdır. Sabit katsayı için test istatistiği,

$$t = \frac{\hat{\alpha} - \alpha}{S_{\hat{\alpha}}}$$

dır. Sabit katsayının sıfıra eşitliği test edildiğinden $\alpha = 0$ alınır ve test istatistiği,

$$t = \frac{\hat{\alpha}}{S_{\hat{\alpha}}}$$

olarak hesaplanır.

Son olarak hipotezlerden birinin kabul kararı hipotez testlerinde açıklandığı gibi verilir. Kısaca tablodan bulunacak değer Z_t ile ifade edilirse, hem tek hem çift taraflı testler ve hem t hem de normal dağılım için $|Z| < |Z_t|$ ise H_0 hipotezi, $|Z| > |Z_t|$ ise H_1 hipotezi kabul edilir.

Sabit katsayının anlamlılığının testinde sıfır hipotezi kabul edilirse ($\alpha = 0$) regresyon modelinin orjinden geçmesi sözkonusudur. Bu sonuç modelin değişkenler arasındaki ilişkiyi iyi açıklamadığı sonucunu doğurur.

Eğim katsayısının testinde hipotezler aynı şekilde oluşturulur. Yani, hipotezlerdeki α 'nın yerini eğim katsayısı β alınır.

Temel Hipotez

$H_0 : \beta = 0$

Karşıt Hipotez

$H_1 : \beta \neq 0$ Çift Taraflı

$\left.\begin{array}{l} H_1 : \beta > 0 \\ H_1 : \beta < 0 \end{array}\right\}$ Tek Taraflı

Eğim katsayısının testinde tablo değerleri sabit katsayının testi gibi belirlenir. Eğim katsayısı için test istatistiği,

$$t = \frac{\hat{\beta} - \beta}{S_{\hat{\beta}}}$$

dir ve $\beta = 0$ olup olmadığı test edildiğinden,

$$t = \frac{\hat{\beta}}{S_{\hat{\beta}}}$$

olacaktır. Test istatistiği hesaplandıktan sonra hangi hipotezin kabul edileceği kararı aynı şekilde verilir. Eğim katsayısının testinde H_0 hipotezinin kabulü X ve Y değişkenleri arasında belirlenen hata payı ile ilişki olmadığını ifade etmektedir.

ÖRNEK: 14 devrelik veri ile basit doğrusal regresyon modeli,

$$Y_i = 3{,}411 - 0{,}473X_i$$
$$(1{,}219) \quad (0{,}135)$$

olarak tahmin edilmiştir.

a) 0,05 hata payı ile sabit katsayının anlamlı olup olmadığını test ediniz.

b) 0,01 hata payı ile eğim katsayısının tek taraflı test ediniz.

a) n = 14

$\hat{\alpha} = 3{,}411$

$S_{\hat{\alpha}} = 1{,}219$

$H_0 : \alpha = 0$

$H_1 : \alpha \neq 0$

$$\mathrm{t}_{\alpha/2,\mathrm{n}-2} = \mathrm{t}_{0,05/(14-2=12)} = 2{,}179$$

$$t = \frac{\hat{\alpha}}{S_{\hat{\alpha}}} = \frac{3{,}411}{1{,}219} = 2{,}798$$

$$|t| > |t_{\alpha/2,n-2}|$$

2,798 >2,179 olduğundan H_1 hipotezi kabul, H_0 hipotezi reddedilir.

b) n = 14

$\hat{\beta} = -0{,}473$

$S_{\hat{\beta}} = 0{,}135$

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

$$\mathrm{t}_{\alpha,\mathrm{n}-2} = \mathrm{t}_{0,01/12} = 3{,}055$$

$$t = \frac{\hat{\beta}}{S_{\hat{\beta}}} = \frac{-0{,}473}{0{,}135} = -3{,}503$$

$$|t| > |t_{\alpha,n-2}|$$

3,503 > 3,055 olduğundan H_1 hipotezi kabul, H_0 hipotezi reddedilir.

13.1.5.2. F Testi

Çoklu regresyon için anlamlı olsa da basit regresyon modelleri için de F testi yapılabilir. F testi bağımsız değişkenlerin tümünün katsayılarının istatistiksel olarak anlamlı olup olmadığını test eder. Basit regresyonda tek bağımsız değişken olduğundan basit regresyonda F testinin t testinden farklı bir anlamı yoktur.

F testi basit regresyonun bağımsız değişkeninin anlamlı olup olmadığını test edecektir. Bu nedenle temel hipotez,

$$H_0 : \beta = 0$$

olarak oluşturulacaktır. F testinde tek taraflı test yapılamaz. Karşıt hipotez H_0 hipotezinin doğru olmadığını belirteceğinden,

$$H_1 : \beta \neq 0$$

olarak oluşturulur.

F test istatistiği regresyon tarafından açıklanan değişkenin, açıklanmayan değişmeye oranına eşittir. Test istatistiği,

$$F = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 / 1}{\sum_{i=1}^{n} (Y_i - \hat{Y}_i) / (n-2)}$$

dir. Test istatistiğinin dağılımı F dağılımıdır. Serbestlik dereceleri (1) ve (n-2)'dir. F test istatistiği F tablosundan serbestlik dereceleri $SD_1=1$ ve $SD_2 = n-2$ ve α hata payı ile bulunan tablo değeri ile karşılaştırılır.

$F < F_{\alpha,1,n-2}$ ise H_0 hipotezi,

$F > F_{\alpha,1,n-2}$ ise H_1 hipotezi

kabul edilir.

F test istatistiği belirlilik katsayısı R^2 ile de hesaplanabilir.

$$F = \frac{R^2}{1-R^2} . \frac{n-2}{1} = \frac{(n-2)R^2}{1-R^2}$$

Basit regresyonda eğim katsayısı için hesaplanan t test istatistiğinin karesi F test istatistiğine eşittir.

$$t_\beta^2 = F$$

Bu eşitlik sadece basit regresyon için geçerlidir.

ÖRNEK: 10 devrelik veri ile regresyon modeli

$$Y_i = 2{,}271 + 0{,}371 X_i$$

$$\sum_{i=1}^{n} (\hat{Y} - \overline{Y})^2 = 20{,}6$$

$$\sum_{i=1}^{n}(Y-\hat{Y}_i)^2 = 8{,}71$$

olarak tahmin edilmiştir. 0,05 hata payı ile modeli F testi ile test ediniz.

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

$SD_1 = 1$

SD_2 = n-2 = 10-2=8

$F_{\alpha,1,n-2} = F_{0,05,1,8} = 5{,}32$

$$F = \frac{\sum(\hat{Y}_i - \overline{Y})^2}{\sum(Y-\hat{Y})/(n-2)} = \frac{20{,}6}{8{,}71/8} = 18{,}92$$

18,92 > 5,32 H_1 hipotezi kabul edilir.

Bu örnekte toplam değişme,

$$\sum(Y_i - \overline{Y})^2 = \sum(\hat{Y}_i - \overline{Y})^2 + \sum(Y_i - \hat{Y}_i)^2$$
$$= 20{,}6 + 8{,}71 = 29{,}31$$

olacaktır. Bu sonuçla,

$$R^2 = \frac{\sum(\hat{Y}_i - \overline{Y})^2}{\sum(Y_i - \overline{Y})^2} = \frac{20{,}6}{29{,}31} = 0{,}7028$$

bulunur ve

$$F = \frac{(n-2)R^2}{1-R^2} = \frac{(8)0{,}7028}{1-0{,}7028} = 18{,}92$$

olarak elde edilir. Bu örnekte β için hesaplanacak t test istatistiği

$$t_\beta^2 = F = 18{,}92$$

$$t_\beta = 4{,}3497$$

olacaktır. $\beta = 0{,}371$ olduğundan,

$$t = \frac{\hat{\beta}}{S_{\hat{\beta}}}$$

$$4{,}3497 = \frac{0{,}371}{S_{\hat{\beta}}}$$

$$S_{\hat{\beta}} = 0{,}08529$$

bulunacaktır.

13.1.6. Regresyon Modelinin Tahmin İçin Kullanılması

Regresyon modeli ile bağımsız değişkenin alacağı farklı değerler için bağımlı değişkenin değerleri tahmin edilebilir. Örneğin, bir işletme yöneticisi ürettikleri malın hangi fiyattan kaç adet satılacağını veya geçmiş yılların ciroları ile karları arasındaki ilişkiden yararlanarak gelecek yılın cirosunu belirleyerek karının ne olacağını tahmin etmek isteyebilir.

Regresyon modeli tahmin için kullanılacaksa, bağımsız değişkenin tahmin edilecek değerinin modelde yerine konması gerekecektir. Bağımsız değişkenin X_o değeri için, bağımlı değişkenin tahmini değeri,

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta} X_0$$

olarak elde edilecektir.

Anakütle Y_o değeri ile örnekten tahmin edilen $\hat{Y}_0$ değeri arasındaki fark tahmin hatası olarak adlandırılır. Tahmin hatasının varyansı,

$$S_T^2 = S_e^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum (X_i - \overline{X})^2} \right]$$

formülü ile hesaplanır. Tahmin hatasının varyansının karekökü S_T ise tahminin standart hatasıdır.

ÖRNEK: Daha önce verilen örnekte bağımsız değişkenin 20 ve 25 değerlerini alması halinde bağımlı değişkenin değerlerini tahmin ediniz ve her iki durum için tahminin standart hatasını hesaplayınız.

$Y_i = - 4,3 + 1,03\ X_i$

$S_e^2 = 0,6446$

$\sum (X_i - \overline{X})^2 = 60$

olarak belirlenmiştir.

$X_o = 20$ için $\hat{Y}_0 = - 4,3 + 1,03\ (20) = 16,3$

ve bu durumda tahminin standart hatası,

$$S_T = \sqrt{S_e^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum (X_i - \overline{X})^2} \right]}$$

$$= \sqrt{0,6446 \left[1 + \frac{1}{5} + \frac{(20 - 10)^2}{60} \right]}$$

$$= 1,3593$$

olacaktır.

$X_1 = 25$ için $Y_1 = - 4,3 + 1,03\ (25) = 21,45$

ve bu durumda tahminin standart hatası,

$$S_T = \sqrt{S_e^2\left[1+\frac{1}{n}+\frac{(X_1-\overline{X})^2}{\sum(X_i-\overline{X})^2}\right]}$$

$$= \sqrt{0{,}6446\left[1+\frac{1}{5}+\frac{(25-10)^2}{60}\right]}$$

$$= 1{,}7862$$

olacaktır. Görüldüğü gibi bağımsız değişkenin tahmin edilen değerinin büyümesi ile tahminin standart hatası da büyümektedir.

13.2. KORELASYON KATSAYISI

Regresyon analizinde değişkenler arasındaki ilişki matematiksel bir model ile açıklanmaktaydı. Regresyon modeli değişkenler arasındaki ilişkinin yönünü, bağımsız değişkenin işareti ile belirtmektedir. Ayrıca değişkenlerin karşılıklı değişimlerinin ölçüsü olarak kovaryansın hesaplanabileceğinden daha önce söz edilmişti.

Kovaryans anakütle için,

$$Kov(X,Y) = \sigma_{XY} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{N}$$

formülü ile hesaplanacaktır. Anakütle kovaryansı alınacak bir örnek ile tahmin ediliyorsa, formül,

$$S_{XY} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{n-1}$$

olacaktır. Örnek kovaryansı,

$$Kov(X,Y) = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{n}$$

anakütle kovaryansının sapmasız tahmincisi değildir. Payda n yerine (n - 1) alındığında formül anakütle varyansının sapmasız tahmincisi olmaktadır.

Kovaryans değişkenler arasındaki ilişkinin yönünü vermekte fakat standart bir ölçü olmadığından ilişkinin derecesini belirlemede yeterli olmamaktadır. Ayrıca, kovaryans standart bir ölçü olmadığından serilerin ilişkilerinin karşılaştırılmasında da yetersiz kalmakta ve bu nedenle korelasyon katsayısı hesaplanmaktadır.

Korelasyon katsayısı, kovaryansın değişkenlerin standart sapmalarına bölünmesi ile elde edilen standart bir ölçüdür. Standart ölçü olduğu için değişkenler arasındaki ilişkinin yönünü belirttiği gibi, ilişkinin derecesini de belirtmektedir.

Verilen tanıma göre anakütle korelasyon katsayısı,

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\dfrac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{N}}{\sigma_X \sigma_Y} = \frac{(X_i - \overline{X})(Y_i - \overline{Y})}{N\sigma_X \sigma_Y}$$

olacaktır.

Anakütle korelasyon katsayısı örnekten tahmin edilirken kovaryans ve standart sapmanın sapmasız tahmincileri kullanılarak,

$$r = \frac{S_{XY}}{S_X S_Y}$$

$$S_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

olduğundan, korelasyon katsayısının tanım formülü,

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)S_X S_Y}$$

olarak elde edilecektir. Burada,

$$S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

ve

$$S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}}$$

formülleri ile hesaplanacaktır.

Korelasyon katsayısının hesaplanmasında regresyonda olduğu gibi bağımlı ve bağımsız değişken ayrımı yoktur. Bu nedenle sadece korelasyon katsayısının hesaplanacağı durumlarda değişkenlerden herhangi biri X_i, diğeri Y_i olarak alınabilir.

Korelasyon katsayısının değeri $\mp$ 1 arasında değişir.

Korelasyon katsayısı şöyle yorumlanacaktır.

$-1 \leq r \leq +1$

r =0 ise değişkenler arasında ilişki yoktur.

r < 0 ise değişkenler arasında ters yönlü ilişki vardır.

r > 0 ise değişkenler arasında doğru yönlü ilişki vardır.

r'nin değeri + 1'e yaklaştıkça doğru yönlü ilişki, -1'e yaklaştıkça ters yönlü ilişki kuvvetlenecektir. r'nin değeri sıfıra yaklaşıyorsa ilişki zayıflayacaktır.

r'nin değeri 0,50 civarında ise ilişkinin orta dereceli olduğu söylenebilir.

Örneğin; r = - 0,85 ise değişkenler arasında r < 0 olduğundan ters yönlü, fakat r'nin değeri 1'e yakın olduğundan kuvvetli ilişki olacaktır. Bu sonucu kısaca, değişkenler arasında ters yönlü kuvvetli ilişki vardır şeklinde yorumlayabiliriz. Eğer r = 0,15 ise, değişkenler arasında r > 0 olduğundan doğru yönlü, fakat r'nin değeri 0'a yakın olduğundan zayıf ilişki olacaktır. Bu sonucu kısaca, değişkenler arasında doğru yönlü zayıf ilişki vardır şeklinde yorumlayabiliriz. r'nin değeri sıfıra ne kadar yakın ise ilişkinin zayıf veya bire ne kadar yakın ise ilişkinin kuvvetli olacağı araştırma konusuna ve araştırmacıya göre değişebilir. Bu nedenle kesin sınırların çizilmesi doğru olmayacaktır. Ayrıca bu bölümde verilecek formüller ile hesaplanacak korelasyon katsayısı ile doğrusal ilişkinin yönü ve derecesi belirlendiğinden, yapılacak yorumların tümü doğrusal ilişki ile ilgili olacaktır. Bu şekilde hesaplanan korelasyon katsayısının değerinin küçük olması, değişkenler arasındaki doğrusal ilişkinin zayıf olduğunu gösterecektir.

ÖRNEK: Aşağıda verilen serinin korelasyon katsayısını tanım formülü ile hesaplayınız.

Y_i	X_i	$Y_i-\overline{Y}$	$X_i-\overline{X}$	$(X_i-\overline{X})(Y_i-\overline{Y})$	$(Y_i-\overline{Y})^2$	$(X_i-\overline{X})^2$
2	5	-3	-2	6	9	4
5	4	0	-3	0	0	9
3	7	-2	0	0	4	0
7	8	2	1	2	4	1
8	11	3	4	12	9	16
25	35			20	26	30

$$r=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{(n-1)S_X S_Y}$$

$$\overline{X}=\frac{\sum X_i}{n}=\frac{35}{5}=7$$

$$\overline{Y}=\frac{\sum Y_i}{n}=\frac{25}{5}=5$$

$$S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}} = \sqrt{\frac{30}{5-1}} = \sqrt{7{,}5} = 2{,}7386$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}} = \sqrt{\frac{26}{5-1}} = \sqrt{6{,}5} = 2{,}5495$$

$$r = \frac{20}{(5-1)(2{,}7386)(2{,}5495)} = \frac{20}{27{,}9282} = 0{,}7161$$

tanım formülünde S_x ve S_y formüllerini yerine koyarak, formülde gerekli kısaltmalar yapılırsa,

$$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1} \cdot \frac{\sum (Y_i - \overline{Y})^2}{n-1}}}$$

korelasyon katsayısının ortalamalardan farklar formülü,

$$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}$$

olarak elde edilecektir. Ortalamadan farklar serileri,

$$x_i = X_i - \overline{X}$$

$$y_i = Y_i - \overline{Y}$$

olarak ifade edilirse formül,

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

şeklinde yazılacaktır.

ÖRNEK: Daha önce verilen örnek için korelasyon katsayısını ortalamadan farklar formülü ile hesaplayınız.

Daha önce,

$$\sum (X_i - \overline{X})(Y_i - \overline{Y}) = 20$$

$$\sum (X_i - \overline{X})^2 = 30$$

$$\sum (Y_i - \overline{Y})^2 = 26$$

olarak hesaplanmıştı.

$$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}$$

$$r = \frac{20}{\sqrt{26{,}30}} = \frac{20}{27{,}9284} = 0{,}7161$$

Ortalamadan farkların gerçek değerleri ile hesaplanması için aşağıda verilen formüllerin kullanılabileceğini daha önce göstermiştik.

$$\sum (X_i - \overline{X})(Y_i - \overline{Y}) = \sum X_i Y_i - n\overline{X}\overline{Y}$$

$$\sum (X_i - \overline{X})^2 = \sum X_i^2 - n\overline{X}^2$$

$$\sum (Y_i - \overline{Y})^2 = \sum Y_i^2 - n\overline{Y}^2$$

Bu formüllerin, ortalamadan farklar ile hesaplanan korelasyon katsayısı formülünde yerine konması ile korelasyon katsayısının gerçek değerlerden hesaplanma formülü,

$$\mathrm{r} = \frac{\sum \mathrm{X_i Y_i} - \mathrm{n}\overline{\mathrm{X}}\overline{\mathrm{Y}}}{\sqrt{[\sum \mathrm{X_i^2} - \mathrm{n}(\overline{\mathrm{X}})^2][\sum \mathrm{Y_i^2} - \mathrm{n}(\overline{\mathrm{Y}})^2]}}$$

olarak elde edilecektir.

ÖRNEK: Daha önce verilen örnek için korelasyon katsayısını gerçek değerler formülü ile hesaplayınız.

Y_i	X_i	$Y_i X_i$	Y_i^2	X_i^2
2	5	10	4	25
5	4	20	25	16
3	7	21	9	49
7	8	56	49	64
8	11	88	64	121
25	35	195	151	275

$$\overline{X} = 7$$

$$\overline{Y} = 5$$

$$r = \frac{\sum XY - n\overline{X}\overline{Y}}{\sqrt{[\sum X_i^2 - n(\overline{X})^2][\sum Y_i^2 - n(\overline{Y})^2]}}$$

$$= \frac{195-(5)(7)(5)}{\sqrt{[275-5(7)^2][151-5(5)^2]}}$$

$$= \frac{20}{\sqrt{30.26}} = \frac{20}{27,9284} = 0,7161$$

Korelasyon katsayısı için standart hata

$$\sigma_r = \sqrt{\frac{1-r^2}{n-2}}$$

formülü ile hesaplanacaktır.

ÖRNEK: Daha önce verilen örnekte n = 5 ve r = 0,71 olarak hesaplanmıştı. Bu örnek için,

$$\sigma_r = \sqrt{\frac{1-(0,71)^2}{5-2}} = 0,40$$

olacaktır.

Korelasyon Katsayısının Testi

Korelasyon katsayısı r, anakütle korelasyon katsayısı ρ 'nun tahmincisidir. Örnekten anakütle korelasyon katsayısı ρ = r olarak tahmin edilir. Korelasyon katsayısı için tahmin hatası,

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

dir.

Tahmin edilen korelasyon katsayısının istatistiksel anlamlılığı için hipotez testi yapılabilir. Temel hipotez,

$$H_0 : \rho = 0$$

şeklinde oluşturulur. Karşıt hipotez ise testin çift veya tek taraflı olmasına göre,

$$H_1 : \rho \neq 0$$

veya

$$H_1 : \rho > 0$$

$$H_1 : \rho < 0$$

olacaktır.

Korelasyon katsayısının anlamlılığının testi için test istatistiği,

$$Z = \frac{r-\rho}{\sigma_r}$$

temel hipotez ρ =0 olduğundan, σ_r 'nin de yerine konması ile,

$$Z = \frac{r}{\sigma_r} = \frac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

olacaktır.

Test istatistiğinin dağılımı n'e bağlı olarak normal dağılım veya (n-2) serbestlik dereceli t dağılımıdır. Hipotezlerden hangisinin kabul edileceği kararı t veya normal dağılım tablolarından bulunacak değerler ile test istatistiği karşılaştırılarak verilecektir. Uygun tablo değerini Z_t ile ifade edersek,

$|Z| > |Z_t|$ olması durumunda H_1 hipotezi,

$|Z| < |Z_t|$ olması durumunda H_0 hipotezi kabul edilecektir.

H_0 hipotezinin, yani ρ =0 olmasının kabulü korelasyon katsayısının istatistiksel olarak anlamsız olduğunu ortaya koyacaktır.

ÖRNEK: Daha önce verilen örnekte n=5'ti ve r =0.71 olarak hesaplanmıştı. 0,05 hata payı ile değişkenler arasındaki doğru yönlü ilişki istatistiksel olarak anlamlı mıdır, inceleyiniz.

$$H_0 : \rho = 0$$

$$H_1 : \rho > 0$$

$t_{\alpha,n-2} = t_{0,05,3} = 2,353$

$$Z = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,71\sqrt{5-2}}{\sqrt{1-(0,71)^2}} = 1,746$$

1,746 < 2,353 olduğundan H_0 hipotezi kabul edilir.

13.3. BASİT DOĞRUSAL REGRESYON VE KORELASYON ARASINDAKİ İLİŞKİ

Basit doğrusal regresyon modelinin bağımsız değişkeninin katsayısı ortalamadan farklar ile,

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

olarak hesaplanıyordu. Pay ve paydayı (n-1)'e bölersek,

$$\hat{\beta} = \frac{\dfrac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}}{\dfrac{\sum_{i=1}(X_i - \overline{X})^2}{n-1}}$$

$$S_{XY}^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

$$S_X^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

olduğundan,

$$\hat{\beta} = \frac{S_{YX}}{S_X^2}$$

olacaktır. Buradan,

$$S_{YX} = \hat{\beta} S_X^2$$

olarak elde edilir. Korelasyon katsayısı,

$$r = \frac{S_{YX}}{S_X S_Y}$$

olduğundan yukarıda yazılan eşitlikle,

$$r = \frac{\hat{\beta} S_X^2}{S_X S_Y} = \hat{\beta}\frac{S_X}{S_Y}$$

ve

$$\hat{\beta} = r\frac{S_Y}{S_X}$$

olarak elde edilir. Görüldüğü gibi S_x ve S_Y ile birlikte $\hat{\beta}$ biliniyorsa r, r biliniyorsa $\hat{\beta}$ hesaplanabilir.

ÖRNEK: Aşağıda verilen seride,

$$\hat{\beta} = r\frac{S_Y}{S_X}$$

olduğunu gösteriniz.

Y_i	X_i	$Y_i-\overline{Y}$	$X_i-\overline{X}$	$(X_i-\overline{X})(Y_i-\overline{Y})$	$(Y_i-\overline{Y})^2$	$(X_i-\overline{X})^2$
5	8	-5	-7	35	25	49
8	7	-2	-8	16	4	64
6	12	-4	-3	12	16	9
10	15	0	0	0	0	0
14	23	4	8	32	16	64
17	25	7	10	70	49	100
60	90			165	110	286

$$\overline{X}=\frac{\sum X_i}{n}=\frac{90}{6}=15$$

$$\overline{Y}=\frac{\sum Y_i}{n}=\frac{60}{6}=10$$

$$Y_i=\hat{\alpha}+\hat{\beta}X_i$$

$$\hat{\beta}=\frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{\sum(X_i-\overline{X})^2}=\frac{165}{286}=0{,}57$$

$$r=\frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum(X_i-\overline{X})^2\sum(Y_i-\overline{Y})^2}}$$

$$r=\frac{165}{\sqrt{286.110}}=\frac{165}{177{,}3696}=0{,}93$$

$$S_X=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}}=\sqrt{\frac{286}{6-1}}=\sqrt{57{,}2}=7{,}56$$

$$S_Y=\sqrt{\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{n-1}}=\sqrt{\frac{110}{6-1}}=\sqrt{22}=4{,}69$$

$$\hat{\beta}=r\frac{S_Y}{S_X}$$

$$=0{,}93\frac{4{,}69}{7{,}56}=0{,}57$$

Regresyon modelinde X_i değerlerini yerine koyarak teorik değerler elde ediliyordu.

$$Y_i=\hat{\alpha}+\hat{\beta}X_i$$

eşitliğinde bağımsız değişkenin katsayısı,

$$\hat{\beta} = r\frac{S_Y}{S_X}$$

ve sabit katsayı,

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= \overline{Y} - r\frac{S_Y}{S_X}\overline{X}$$

olduğundan, eşitlikleri yerine koyarsak,

$$\hat{Y}_i = \overline{Y} - r\frac{S_X}{S_Y}\overline{X} + r\frac{S_Y}{S_X}X_i$$

$$= \overline{Y} + r\frac{S_X}{S_Y}(X_i - \overline{X})$$

olacak ve bu şekilde tahminde yapılabilecektir.

ÖRNEK: Daha önce verilen örnekte $X_i = 25$ için teorik değeri hesaplayarak,

$$\hat{Y}_i = \overline{Y} + r\frac{S_Y}{S_X}(X_i - \overline{X})$$

olduğunu gösteriniz.

$$\overline{X} = 15$$

$$\overline{Y} = 10$$

$$\hat{\beta} = 0{,}57$$

olduğundan,

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= 10 - 0{,}57\ (15) = 1{,}45$$

$$\hat{Y} = 1{,}45 + 0{,}57(25) = 15{,}7$$

bulunur.

$$\hat{Y}_i = \overline{Y} + r\frac{S_Y}{S_X}(X_i - \overline{X})$$

$r = 0{,}93$

$S_Y = 4{,}69$

$S_x = 7{,}56$

olduğundan

$$\hat{Y} = 10 + 0{,}93\frac{4{,}69}{7{,}56}(25 - 15) = 15{,}7$$

olarak elde edilir.

Belirlilik katsayısı,

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

ve

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 - \hat{\beta}\sum_{i01}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$$

olduğundan belirlilik katsayısı,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 - \hat{\beta}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

olarak yazılabilir. Gerekli kısaltmalar yapılırsa,

$$R^2 = 1 - 1 + \frac{\hat{\beta}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

pay ve payda (n-1)'e bölünürse,

$$R^2 = \frac{\hat{\beta}S_{YX}}{S_Y^2}$$

olacaktır. Korelasyon katsayısı,

$$r = \frac{S_{YX}}{S_Y S_X}$$

olduğundan,

$$S_{YX} = rS_Y S_X$$

olarak elde edilir. Bu eşitlik yukarıda verilen R^2 eşitliğinde yerine konursa,

$$R^2 = \frac{\hat{\beta}rS_Y S_X}{S_Y^2} = r\hat{\beta}\frac{S_X}{S_Y}$$

ve

$$r = \hat{\beta}\frac{S_X}{S_Y}$$

olduğundan,

$$R^2 = r.r = r^2$$

olarak bulunur. Görüldüğü gibi basit doğrusal regresyonda korelasyon katsayısının karesi belirlilik katsayısına eşittir.

ÖRNEK: Daha önce verilen örnekte

$$r^2 = R^2$$

olduğunu gösteriniz.

Örnekte r = 0,93 , regresyon modeli,

$Y_i = 1,45 + 0,57\ X_i$

$$\sum (X_i - \overline{X})(Y_i - \overline{Y}) = 165$$

$$\sum (X_i - \overline{X})^2 = 110$$

olarak belirlenmişti.

$$\sum (Y_i - \hat{Y})^2 = \sum (Y_i - \overline{Y})^2 - \hat{\beta}\sum (X_i - \overline{X})(Y_i - \overline{Y})$$

$$= 110 - 0,57\ (165) = 15,95$$

$$R^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \overline{Y})^2} = 1 - \frac{15,95}{110}$$

$$= 1 - 0,145 = 0,855$$

$$r^2 = (0,93)^2 = 0,8649$$

olarak bulunur. Hesaplama şekilleri arasındaki fark değerlerin yaklaşık alınmasından kaynaklanmaktadır.

Regresyon modelinde X_i bağımsız, Y_i ise bağımlı değişken olarak alınmaktaydı. X_i bağımlı, Y_i bağımsız değişken alınarak da regresyon modeli oluşturulabilir. X_i'nin bağımsız, Y_i'nin bağımlı değişken olduğu regresyon modeli,

$$Y_i = \hat{\alpha}_{XY} - \hat{\beta}_{YX} X_i$$

ve X_i'nin bağımlı, Y_i'nin bağımsız değişken olduğu regresyon modeli,

$$X_i = \hat{\alpha}_{XY} + \hat{\beta}_{YX} Y_i$$

şeklinde ifade edilirse, ikinci model ile ilgili hesaplamaların yapılabilmesi için daha önce verilen formüllerde X_i ile Y_i'nin yer değiştirmesi yeterli olacaktır.

Yukarıda verilen modellerin bağımsız değişkenlerinin katsayıları,

$$\hat{\beta}_{YX} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

$$\hat{\beta}_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

formülleri ile hesaplanacaktır. Verilen bu katsayılar çarpılarak kare kökleri alınırsa,

$$\sqrt{\hat{\beta}_{YX}\hat{\beta}_{XY}} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} . \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

$$= \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}} = r$$

olarak elde edilir. Bu durumda β_{YX} ve β_{XY}'nin işaretleri birbiri ile aynı olacak ve korelasyon katsayısı, regresyon modelinin katsayıları ile aynı işareti taşıyacaktır.

ÖRNEK: Aşağıda verilen seri için,

$$r = \sqrt{\hat{\beta}_{YX}\hat{\beta}_{XY}}$$

olduğunu gösteriniz.

Y_i	X_i	$Y_i - \overline{Y}$	$X_i - \overline{X}$	$(X_i - \overline{X})(Y_i - \overline{Y})$	$(Y_i - \overline{Y})^2$	$(X_i - \overline{X})^2$
2	17	-3	5	-15	9	25
4	14	-1	2	-2	1	4
3	11	-2	-1	2	4	1
6	10	1	-2	-2	1	4
10	8	5	-4	-20	25	16
25	60			-37	40	50

$$\overline{X} = \frac{\sum X_i}{n} = \frac{60}{5} = 12$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{25}{5} = 5$$

$$\hat{\beta}_{YX} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

$$= \frac{-37}{50} = -0{,}74$$

$$\hat{\beta}_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

$$= \frac{-37}{40} = -0{,}925$$

$$r = \sqrt{\hat{\beta}_{YX}\hat{\beta}_{XY}}$$

$$= \sqrt{(-0{,}74)(-0{,}925)} = -0{,}827$$

$$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}$$

$$= \frac{-37}{\sqrt{50.40}} = \frac{-37}{44{,}721} = -0{,}827$$

13.4. EĞRİSEL KORELASYON VE REGRESYON

Buraya kadar yapılan açıklamalarda iki değişken arasındaki ilişkinin doğrusal olduğu kabul ediliyordu. Değişkenler arasındaki ilişkinin doğrusal olmadığı durumlarda, bu ilişkiyi açıklayacak eğrisel modelin belirlenmesi gerekecektir.

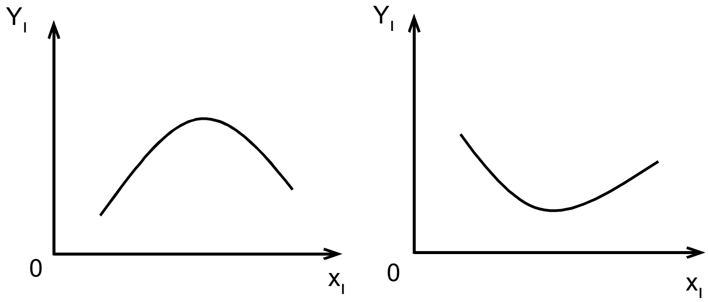
Değişkenler arasındaki eğrisel ilişkiyi açıklamak için kullanılacak modelin matematiksel şeklini belirlemek için yapılabilecek en basit işlem serinin grafiğini çizmektir. Çizilen grafik yardımı ile model belirlenebilirse de, grafikler yanıltıcı sonuçlar verebilirler. Bu nedenle, ilişkiyi açıklamak için uygun olabilecek modellerin denenerek, bunlardan ilişkiyi en iyi açıklayanının, yani belirlilik katsayısı en yüksek olanının seçilmesi daha doğru olacaktır.

Değişkenler arasındaki ilişkiyi açıklamak için çeşitli eğrisel modeller kullanılabilir. Burada en çok kullanılan birkaç eğrisel modele yer verilecektir.

Serinin tek maksimumu veya minimumu varsa değişkenler arasındaki ilişki,

$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$

modeli ile açıklanabilir.



Şekil 13.3

Yukarıda verilen anakütle regresyon modeli alınacak örnekten,

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \varepsilon_i$$

olarak tahmin edilecektir. Modelin parametrelerinin en küçük kareler yöntemi ile tahmini, doğrusal modelin parametrelerinin tahminine benzer şekilde yapılarak model için normal denklemler,

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2$$

$$\sum_{i=1}^{n} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^3$$

$$\sum_{i=1}^{n} Y_i X_i^2 = \hat{\beta}_0 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} X_i^3 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^4$$

olarak elde edilir. Bu durumda hata terimlerinin kareleri toplamı,

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} Y_i^2 - \hat{\beta}_0 \sum_{i=1}^{n} Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i Y_i - \hat{\beta}_2 \sum_{i=1}^{n} X_i^2 Y_i$$

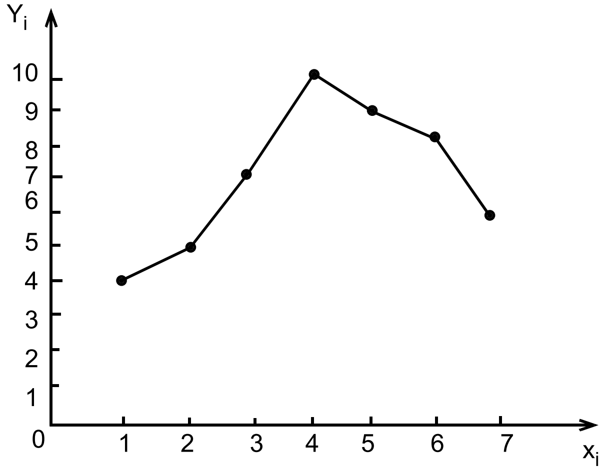
formülü ile hesaplanacaktır. Korelasyon katsayısı ise daha sonra çoklu regresyonda yapılacak açıklamalara göre belirlenecektir.

ÖRNEK: Aşağıdaki serinin grafiğini çizerek,

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \varepsilon_i$$

regresyon modelinin parametrelerini tahmin ediniz.

Y_i	X_i	X_i^2	X_i^3	X_i^4	$Y_i\ X_i$	$X_i^2\ Y_i$
4	1	1	1	1	4	4
5	2	4	8	16	10	20
7	3	9	27	81	21	63
10	4	16	64	256	40	160
9	5	25	125	625	45	225
8	6	36	216	1296	48	288
6	7	49	343	2401	42	294
49	28	140	784	4676	210	1054



$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2$$

$$\sum_{i=1}^{n} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^3$$

$$\sum_{i=1}^{n} Y_i X_i^2 = \hat{\beta}_0 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} X_i^3 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^4$$

$$49 = 7\hat{\beta}_0 + 28\hat{\beta}_1 + 140\hat{\beta}_2$$
$$210 = 28\hat{\beta}_0 + 140\hat{\beta}_1 + 784\hat{\beta}_2$$
$$1054 = 140\hat{\beta}_0 + 784\hat{\beta}_1 + 4676\hat{\beta}_2$$

$$-4/49 = 7\hat{\beta}_0 + 28\hat{\beta}_1 + 140\hat{\beta}_2$$
$$210 = 28\hat{\beta}_0 + 140\hat{\beta}_1 + 784\hat{\beta}_2$$

$$-196 = -28\hat{\beta}_0 - 112\hat{\beta}_1 - 560\hat{\beta}_2$$
$$210 = 28\hat{\beta}_0 + 140\hat{\beta}_1 + 784\hat{\beta}_2$$

$$14 = 0 + 28\hat{\beta}_1 + 224\hat{\beta}_2$$

$$-20/49 = 7\hat{\beta}_0 + 28\hat{\beta}_1 + 140\hat{\beta}_2$$
$$1054 = 140\hat{\beta}_0 + 784\hat{\beta}_1 + 4676\hat{\beta}_2$$

$$-980 = -140\hat{\beta}_0 - 560\hat{\beta}_1 - 2800\hat{\beta}_2$$
$$1054 = 140\hat{\beta}_0 + 784\hat{\beta}_1 + 4676\hat{\beta}_2$$

$$74 = 0 + 224\hat{\beta}_1 + 1876\hat{\beta}_2$$

$$-8/14 = 28\hat{\beta}_1 + 224\hat{\beta}_2$$
$$74 = 224\hat{\beta}_1 + 1876\hat{\beta}_2$$

$$-112 = -224\hat{\beta}_1 - 1792\hat{\beta}_2$$
$$74 = 224\hat{\beta}_1 + 1876\hat{\beta}_2$$

$$-38 = 0 + 84\hat{\beta}_2$$

$$\beta_2 = \frac{-38}{84} = -0{,}45$$

$$14 = 28\hat{\beta}_1 + 224(-0{,}45)$$

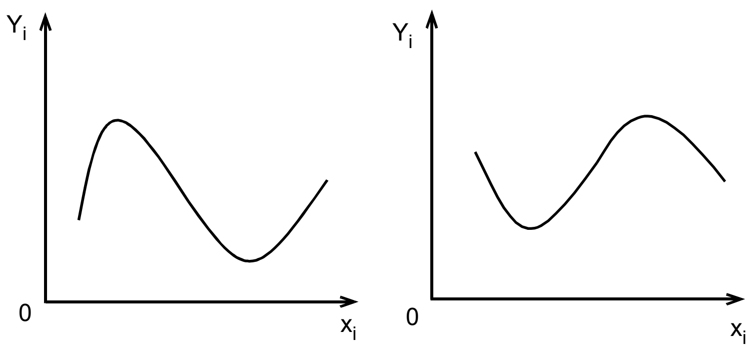
$$\beta_1 = \frac{14 + 100{,}8}{28} = 4{,}1$$

$$49 = 7\hat{\beta}_0 + 28(4{,}1) + 140(-0{,}45)$$

$$\beta_0 = \frac{49 - 114{,}8 + 63}{7} = -0{,}4$$

$$Y_i = -0{,}4 + 4{,}1X_i - 0{,}45X_i^2$$

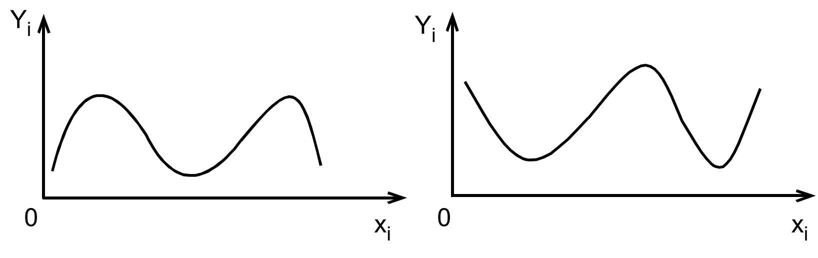
Serinin bir maksimumu ve bir minimumu varsa (Şekil 13.4) örnek regresyon modeli,



Şekil 13.4

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \hat{\beta}_3 X_i^3 + \varepsilon_i$$

serinin iki maksimumu ve bir minimumu veya bir maksimumu ve iki minimumu varsa(Şekil 13.5) örnek regresyon modeli,

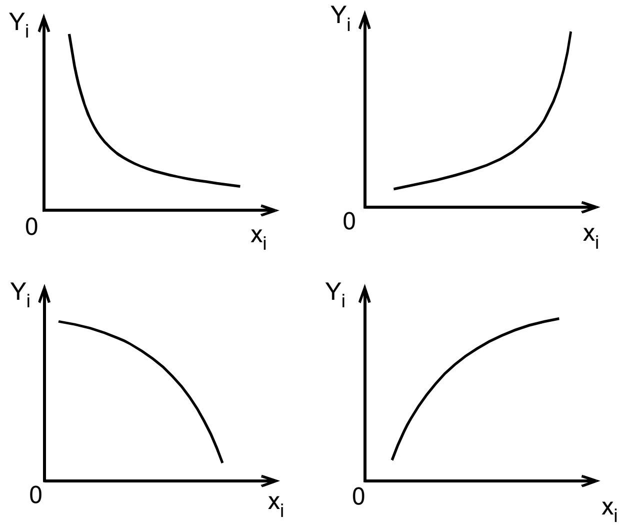


Şekil 13.5

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \hat{\beta}_3 X_i^3 + \hat{\beta}_4 X_i^4 + \varepsilon_i$$

olacaktır. Görüldüğü gibi denklemde bağımsız değişkenin kuvveti serinin maksimum ve minimum sayıları toplamından bir fazladır. Bu tür modeller için normal denklemler doğrusal modelin normal denklemlerinin elde edilmesine benzer şekilde elde edilecek ve çözüm için tahmin edilecek parametre sayısı kadar denklem olacaktır.

Serinin maksimum ve minimumu yoksa, fakat ilişki doğrusal değilse (Şekil 13.6), değişkenler arasındaki ilişki yarı logaritmik veya tam logaritmik modellerle açıklanabilir.



Şekil 13.6

Değişkenlerden sadece birinin logaritmasının alındığı modeller yarı logaritmik model olarak adlandırılmaktadır.

$$Y_i = \hat{\alpha} + \hat{\beta} \ln X_i + e_i$$

ve

$$\ln Y_i = \hat{\alpha} + \hat{\beta} X_i + e_i$$

modelleri yarı logaritmik modellerdir. Logaritmalar 10 veya e tabanına göre alınabilir. Her iki değişkenin logaritmasının alındığı modele ise tam logaritmik model denilmektedir.

$$Y_i = \hat{\alpha} . X_i^{\hat{\beta}} . e^{e_i}$$

modelin logaritması alınırsa,

$$\ln Y_i = \ln \hat{\alpha} + \hat{\beta} \ln X_i + e_i$$

olacaktır. Logaritmik modellerde $\ln e_i$ olması gereken hata terimleri kolaylık olması için e_i ile ifade edilmiştir.

Yarı logaritmik ve tam logaritmik modellerin normal denklemleri ve korelasyon katsayısı tanım formülleri doğrusal model ile aynı şekilde elde edilirler. Elde edilecek formüller, Xi yerine $\ln X_i$ ve Y_i yerine $\ln Y_i$'nin yer aldığı doğrusal regresyon formülleri ile aynı olduklarından burada formüllerin çıkartılması gösterilmeyecektir. Ayrıca, logaritma alındığında ortalamadan farklar formülleri ile yapılacak işlemler güçleşeceğinden sadece gerçek değerler ile ilgili formüller verilecektir.

$$Y_i = \hat{\alpha} + \hat{\beta} \ln X_i + e_i$$

Yarı logaritmik regresyon modeli için normal denklemler,

$$\sum_{i=1}^{n} Y_i = n\hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} \ln X_i$$

$$\sum_{i=1}^{n} Y_i (\ln X_i) = \hat{\alpha} \sum_{i=1}^{n} \ln X_i + \hat{\beta} \sum_{i=1}^{n} (\ln X_i)^2$$

olacaktır. Bu denklemlerden,

$$\hat{\beta} = \frac{\sum_{i=1}^{n} Y_i (\ln X_i) - n\overline{Y}(\overline{\ln X})}{\sum_{i=1}^{n} (\ln X_i)^2 - n(\overline{\ln X})^2}$$

ve

$$\hat{\alpha} = \overline{Y} - \hat{\beta}(\overline{\ln X})$$

olarak elde edilir. Bu durumda korelasyon katsayısı,

$$r = \frac{\sum_{i=1}^{n} Y_i (\ln X_i) - n\overline{Y}(\overline{\ln X})}{\sqrt{[\sum_{i=1}^{n} (\ln X_i)^2 - n(\overline{\ln X})^2][\sum_{i=1}^{n} Y_i^2 - n(\overline{Y})^2]}}$$

formülü ile hesaplanacaktır. Hata terimlerinin kareleri toplamı,

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} Y_i^2 - \hat{\alpha} \sum_{i=1}^{n} Y_i - \hat{\beta} \sum_{i=1}^{n} Y_i (\ln X_i)$$

olacaktır.

ÖRNEK: Aşağıda verilen seri için $Y_i = \hat{\alpha} + \hat{\beta} \ln X_i + e_i$ regresyon modelinin parametrelerini, S_e^2 değerini, korelasyon katsayısını, belirlilik katsayısını hesaplayınız ve $X_6 = 10$ için $\hat{Y}_6$ değerini tahmin ediniz.

Y_i	X_i	$\ln X_i$	$(\ln X_i)^2$	$Y_i(\ln X_i)$	Y_i^2
1	2	0,69314	0,48044	0,69314	1
4	5	1,60943	2,59026	6,43772	16
7	6	1,79175	3,21036	12,54225	49
13	8	2,07944	4,32407	27,03272	169
15	9	2,19722	4,82777	32,9583	225
40	30	8,37098	15,4329	79,66413	460

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{40}{5} = 8$$

$$(\overline{\ln X}) = \frac{\sum \ln X_i}{n} = \frac{8{,}37098}{5} = 1{,}67419$$

$$\hat{\beta} = \frac{\sum Y_i(\ln X_i) - n\overline{Y}(\overline{\ln X})}{\sum (\ln X_i)^2 - n(\overline{\ln X})^2}$$

$$\hat{\beta} = \frac{79{,}66413 - (5)(8)(1{,}67419)}{15{,}43290 - 5(1{,}67419)^2}$$

$$= \frac{12{,}69653}{1{,}41834} = 8{,}95$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}(\overline{\ln X})$$

$= 8 - 8{,}95(1{,}67419)$

$= 8-14{,}98 = -6{,}98$

$$Y_i = -6{,}98 + 8{,}95 \ln X_i$$

$$r = \frac{\sum Y_i(\ln X_i) - n\overline{Y}(\overline{\ln X})}{\sqrt{[\sum (\ln X_i)^2 - n(\overline{\ln X})^2][\sum Y_i^2 - n(\overline{Y})^2]}}$$

$$r = \frac{79{,}66413 - 5(8)(1{,}67413)}{\sqrt{[15{,}43290 - 5(1{,}67419)^2][460 - 5(8)^2]}}$$

$$= \frac{12{,}69653}{\sqrt{(1{,}41834)(140)}}$$

$$= \frac{12,69653}{14,09140} = 0,90$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} Y_i^2 - \hat{\alpha}\sum_{i=1}^{n} Y_i - \hat{\beta}\sum_{i=1}^{n} Y_i(\ln X_i)$$

= 460 - (-6,98) (40) - 8,95 (79,66413)

= 460 + 279,2 - 712,99

= 26,21

$$S_e^2 = \frac{\sum e_i^2}{n-2}$$

$$= \frac{26{,}21}{5-2} = 8{,}73$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \overline{Y})^2}$$

$$= 1 - \frac{26{,}21}{140} = 1 - 0{,}187 = 0{,}813$$

$\hat{Y}_6$ = -6,98 + 8,95 ($\ln X_6$)

= -6,98 + 8,95 (ln10)

= -6,98 + 8,95 (2,30258)

= 13,62

Yarı logaritmik regresyon modeli,

$$\ln Y_i = \hat{\alpha} + \hat{\beta} X_i$$

ise normal denklemler,

$$\sum_{i=1}^{n} (\ln Y_i) = n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i(\ln Y_i) = \hat{\alpha}\sum_{i=1}^{n} X_i + \hat{\beta}\sum_{i=1}^{n} X_i^2$$

olacaktır. Bu denklemlerden,

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i(\ln Y_i) - n(\overline{X})(\overline{\ln Y})}{\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2}$$

ve

$$\hat{\alpha} = (\overline{\ln Y}) - \hat{\beta}\overline{X}$$

olarak elde edilir. Bu durumda korelasyon katsayısı,

$$r = \frac{\sum_{i=1}^{n} X_i(\ln Y_i) - n\overline{X}(\overline{\ln Y})}{\sqrt{[\sum_{i=1}^{n} X_i^{\,2} - n(\overline{X})^2][\sum_{i=1}^{n} (\ln Y_i)^2 - n(\overline{\ln Y})^2]}}$$

formülü ile hesaplanacaktır. Hata terimlerinin kareleri toplamı ise,

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \ln Y_i^2 - \hat{\alpha}\sum_{i=1}^{n}(\ln Y_i) - \hat{\beta}\sum_{i=1}^{n} X_i(\ln Y_i)$$

olacaktır.

ÖRNEK: Daha önce verilen seri için $\ln Y_i = \hat{\alpha} + \hat{\beta}X_i$ regresyon modelinin parametrelerini, S_e^2 değerini, korelasyon katsayısını, belirlilik katsayısını hesaplayınız ve $X_6 = 10$ için $\hat{Y}_6$ değerini tahmin ediniz.

Y_i	X_i	$\ln Y_i$	X_i^2	$X_i(\ln Y_i)$	$(\ln Y_i)^2$
1	2	0	4	0	0
4	5	1,38629	25	6,93145	1,92179
7	6	1,94591	36	11,67546	3,78656
13	8	2,56494	64	20,51952	6,57891
15	9	2,70805	81	24,37245	7,33353
40	30	8,60519	210	63,49888	19,62079

$$\overline{X} = \frac{\sum X_i}{n} = \frac{30}{5} = 6$$

$$(\overline{\ln Y}) = \frac{\sum \ln Y_i}{n} = \frac{8,60519}{5} = 1,72103$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i (\ln Y_i) - n\overline{X}(\overline{\ln Y})}{\sum_{i=1}^{n} X_i^{\,2} - n(\overline{X})^2}$$

$$\hat{\beta} = \frac{63{,}49888 - (5)(6)(1{,}72103)}{210 - 5(6)^2}$$

$$= \frac{11{,}86798}{30} = 0{,}39$$

$$\hat{\alpha} = (\overline{\ln Y}) - \hat{\beta}(\overline{X})$$

= 1,72103 – 0,39(6)

= 1,72103 –2,34 = -0,62

$$\ln\ Y_i = -0{,}62 + 0{,}39 X_i$$

$$r = \frac{\sum X_i (\ln Y_i) - n\overline{X}(\overline{\ln Y})}{\sqrt{[\sum X_i^{\,2} - n(\overline{X})^2][\sum (\ln Y_i)^2 - n(\overline{\ln Y})^2]}}$$

$$r = \frac{63{,}49888 - 5(6)(1{,}72103)}{\sqrt{[210 - 5(6)^2][19{,}62079 - 5(1{,}72103)^2]}}$$

$$= \frac{11{,}86798}{\sqrt{(30)(4{,}81)}} = \frac{11{,}86798}{12{,}0124} = 0{,}98$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (\ln Y_i)^2 - \hat{\alpha}\sum_{i=1}^{n} (\ln Y_i) - \hat{\beta}\sum_{i=1}^{n} X_i (\ln Y_i)$$

= 19,62079 - (-0,62) (8,60519) – 0,39 (63,49888)

= 0,19144

$$S_e^2 = \frac{\sum e_i^2}{n-2}$$

$$= \frac{0{,}19144}{5-2} = 0{,}06381$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum [\ln Y_i - (\overline{\ln Y})]^2}$$

$$\sum[\ln Y_i - (\overline{\ln Y})]^2 = \sum(\ln Y_i)^2 - n(\overline{\ln Y})^2$$

$$= 19{,}62079 - 5(1{,}72103)^2 = 4{,}81106$$

$$R^2 = 1 - \frac{0{,}19144}{4{,}81106}$$

$$=1 - 0{,}039 = 0{,}961$$

$$\ln \hat{Y}_6 = -0{,}62 + 0{,}39X_6$$

$$= -0{,}62 + 0{,}39(10)$$

$$= -0{,}62 + 3{,}9 = 3{,}28$$

$$\hat{Y}_6 = 26{,}57$$

Tam logaritmik regresyon modeli,

$$\ln Y_i = \ln\hat{\alpha} + \hat{\beta}\ln X_i$$

ise normal denklemler,

$$\sum_{i=1}^{n}(\ln Y_i) = n\ln\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n}(\ln X_i)$$

$$\sum_{i=1}^{n}(\ln Y_i)(\ln X_i) = \ln\hat{\alpha}\sum_{i=1}^{n}(\ln X_i) + \hat{\beta}\sum_{i=1}^{n}(\ln X_i)^2$$

olacaktır. Bu denklemlerden,

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(\ln X_i)(\ln Y_i) - n(\overline{\ln X})(\overline{\ln Y})}{\sum_{i=1}^{n}(\ln X)^2 - n(\overline{\ln X})^2}$$

ve

$$\ln\hat{\alpha} = (\overline{\ln Y}) - \hat{\beta}(\overline{\ln X})$$

olarak elde edilir. Bu durumda korelasyon katsayısı,

$$r = \frac{\sum_{i=1}^{n}(\ln X_i)(\ln Y_i) - n(\overline{\ln X})(\overline{\ln Y})}{\sqrt{[\sum_{i=1}^{n}(\ln X_i)^2 - n(\overline{\ln X})^2][\sum_{i=1}^{n}(\ln Y_i)^2 - n(\overline{\ln Y})^2]}}$$

formülü ile hesaplanacaktır. Hata terimlerinin kareleri toplamı ise,

$$\sum_{i=1}^{n}e_i^2 = \sum_{i=1}^{n}\ln Y_i^2 - (\ln\hat{\alpha})\sum_{i=1}^{n}\ln Y_i - \hat{\beta}\sum_{i=1}^{n}(\ln X_i)(\ln Y_i)$$

olacaktır.

ÖRNEK: Daha önce verdiğimiz $\ln Y_i = \ln\hat{\alpha} + \hat{\beta}\ln X_i$ regresyon modelinin parametrelerini, S_e^2 değerini, korelasyon katsayısını, belirlilik katsayısını hesaplayınız ve $\hat{X}_6 = 10$ için Y_6 değerini tahmin ediniz.

Y_i	X_i	$\ln Y_i$	$\ln X_i$	$(\ln X_i)(\ln Y_i)$
1	2	0	0,69314	0
4	5	1,38629	1,60943	2,23113
7	6	1,94591	1,79175	3,48658
13	8	2,56494	2,07944	5,33363
15	9	2,70805	2,19722	5,95018
40	30	8,60519	8,37098	17,00152

Daha önce,

$(\overline{\ln Y}) = 1,72103 \quad \sum(\ln Y_i)^2 = 19,62079$

$(\overline{\ln X}) = 1,67419 \quad \sum(\ln X_i)^2 = 15,43290$

olarak hesaplanmıştı.

$$\hat{\beta} = \frac{\sum(\ln X_i)(\ln Y_i) - n(\overline{\ln X})(\overline{\ln Y})}{\sum(\ln X_i)^2 - n(\overline{\ln X})^2}$$

$$= \frac{17,00152 - 5(1,67419)(1,72103)}{15,43290 - 5(1,67419)^2}$$

$$= \frac{2,59486}{1,41834} = 1,82$$

$$\ln\hat{\alpha} = (\overline{\ln Y}) - \hat{\beta}(\overline{\ln X})$$

$= 1,72103 - 1,82(1,67419)$

$= 1,72103 - 3,04702$

$= -1,32$

$$\ln Y_i = -1,32 + 1,82(\ln X_i)$$

$$r = \frac{\sum(\ln X_i)(\ln Y_i) - n(\overline{\ln X})(\overline{\ln Y})}{\sqrt{[\sum(\ln X_i)^2 - n(\overline{\ln X})^2][\sum(\ln Y_i)^2 - n(\overline{\ln Y})^2]}}$$

$$= \frac{17,00152 - 5(1,67419)(1,72103)}{\sqrt{[15,43290 - 5(1,67419)^2][19,62079 - 5(1,72103)^2}}$$

$$= \frac{2,59486}{\sqrt{1,41834.4,81106}}$$

$$= \frac{2,59486}{2,61222} = 0,99$$

$$\sum e_i^2 = \sum (\ln Y_i)^2 - (\ln \hat{\alpha})\sum \ln Y_i - \hat{\beta}\sum (\ln X_i)(\ln Y_i)$$

$$= 19,62079 - (-1,32)(8,60519) - 1,82(17,00152)$$

$$= 0,03687$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum [\ln Y_i - (\overline{\ln Y})]^2}$$

$$= 1 - \frac{0,03687}{4,81106}$$

$$=1 - 0,007 = 0,993$$

$$S_e^2 = \frac{\sum e_i^2}{n-2}$$

$$= \frac{0,03687}{5-2} = 0,01229$$

$$\ln \hat{Y}_6 = -1,32 + 1,82(\ln X_6)$$

$$= -1,32 + 1,82(2,30258)$$

$$= 2,87069$$

$$\hat{Y}_6 = 17,64$$

Aynı değişkenler için farklı matematiksel yapıdaki modeller denendiğinde bunlardan hangisinin değişkenler arasındaki ilişkiyi daha iyi açıkladığının belirlenmesi gerekecektir. Modellerden hata terimlerinin kareleri toplamı veya varyansı küçük olan, değişkenler arasındaki ilişkiyi daha iyi açıklamaktadır. Modellerin hata terimlerinin varyansı veya karelerinin toplamı ile karşılaştırılabilmesi için modellerin bağımlı değişkenlerinin aynı yapıda olması gerekmektedir. Bu durumda hata terimlerinin varyansı veya kareleri toplamı daha küçük olan model değişkenler arasındaki ilişkiyi daha iyi açıklayacaktır.

Uygun modelin belirlenmesinde belirlilik katsayısı da kullanılabilir. Bağımlı değişkenler aynı yapıda olmadığında da modellerin belirlilik katsayısı ile karşılaştırılması mümkündür. Bağımlı değişkenlerin bazıları gerçek değerlerle, bazıları logaritmik olarak ifade ediliyorsa bu modeller hata terimlerinin tahmincilerinin varyansları veya karelerinin toplamı ile karşılaştırılamazlar.

Bu durumda belirlilik katsayısı büyük olan model, değişkenler arasındaki ilişkiyi daha iyi açıklayacağından bu modelin seçilmesi uygun olacaktır.

ÖRNEK: Daha önce verilen örnek için tam logaritmik ve yarı logaritmik modeller belirlenmişti. Aynı seri için,

$$Y_i = \hat{\alpha} + \hat{\beta} X_i$$

ve

$$Y_i = \hat{\alpha} + \hat{\beta} X_i^2$$

Modellerinin katsayılarını tahmin ediniz.

Y_i	X_i	X_i^2	X_iY_i	Y_i^2	$X_i^2Y_i$	X_i^3	X_i^4
1	2	4	2	1	4	8	16
4	5	25	20	16	100	125	625
7	6	36	42	49	252	216	1296
13	8	64	104	169	832	512	4096
15	9	81	135	225	1215	729	5661
40	30	210	303	460	2403	1590	12594

$$\overline{X} = 6$$

$$\overline{Y} = 8$$

$$\hat{\beta} = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

$$= \frac{303 - 5(6)(8)}{210 - 5(6)^2}$$

$$= \frac{63}{30} = 2{,}1$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= 8 - 2{,}1(6) = -4{,}6$$

$$Y_i = -4{,}6 + 2{,}1 X_i$$

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$$

$$= 460 - (-4{,}6)(40) - 2{,}1(303) \quad = 7{,}7$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y - \overline{Y})^2}$$

$$= 1 - \frac{\sum e_i^2}{\sum Y^2 - n\overline{Y}^2}$$

$$= 1 - \frac{7{,}7}{140}$$

$$= 1 - 0{,}055 = 0{,}945$$

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2$$

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

$$\sum X_i Y_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 + \hat{\beta}_2 \sum X_i^3$$

$$\sum X_i^2 Y_i = \hat{\beta}_0 \sum X_i^2 + \hat{\beta}_1 \sum X_i^3 + \hat{\beta}_2 \sum X_i^4$$

$$-6/40 = 5\hat{\beta}_0 + 30\hat{\beta}_1 + 210\hat{\beta}_2$$

$$303 = 30\hat{\beta}_0 + 210\hat{\beta}_1 + 1590\hat{\beta}_2$$

$$-240 = -30\hat{\beta}_0 - 180\hat{\beta}_1 - 1260\hat{\beta}_2$$

$$303 = 30\hat{\beta}_0 + 210\hat{\beta}_1 + 1590\hat{\beta}_2$$

$$63 = 0 + 30\hat{\beta}_1 + 330\hat{\beta}_2$$

$$-42/40 = 5\hat{\beta}_0 + 30\hat{\beta}_1 + 210\hat{\beta}_2$$

$$2403 = 210\hat{\beta}_0 + 1590\hat{\beta}_1 + 12594\hat{\beta}_2$$

$$-1680 = -210\hat{\beta}_0 - 1260\hat{\beta}_1 - 8820\hat{\beta}_2$$

$$2403 = 210\hat{\beta}_0 + 1590\hat{\beta}_1 + 12594\hat{\beta}_2$$

$$723 = 0 + 330\hat{\beta}_1 + 3774\hat{\beta}_2$$

$$-11/63 = 30\hat{\beta}_1 + 330\hat{\beta}_2$$

$$723 = 330\hat{\beta}_1 + 3774\hat{\beta}_2$$

$$-693 = -330\hat{\beta}_1 - 3630\hat{\beta}_2$$

$$723 = 330\hat{\beta}_1 + 3774\hat{\beta}_2$$

$$30 = 144\hat{\beta}_2$$

$$\hat{\beta}_2 = \frac{30}{144} = 0{,}20$$

$$63 = 30\hat{\beta}_1 + 330(0{,}20)$$

$$\hat{\beta}_1 = \frac{63-66}{30}$$

$$= \frac{-3}{30} = -0{,}10$$

$$40 = 5\,\hat{\beta}_0 + 30(-0{,}10) + 210(0{,}20)$$

$$\hat{\beta}_0 = \frac{40+3-42}{5}$$

$$= \frac{1}{5} = 0{,}2$$

$$Y_i = 0{,}2 - 0{,}1X_i + 0{,}2X_i^2$$

$$\sum e_i^2 = \sum Y^2 - \hat{\beta}_0\sum Y_i - \hat{\beta}_1\sum X_iY_i - \hat{\beta}_2\sum X_i^2Y_i$$

$$= 460 - 0{,}2(40) - (-0{,}1)(303) - 0{,}2(2403) = 1{,}7$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum Y_i^2 - n\overline{Y}^2}$$

$$= 1 - \frac{1{,}7}{460 - 5(8)^2}$$

$$= 1 - \frac{1{,}7}{140}$$

$$= 1 - 0{,}012 \quad = 0{,}988$$

13.5. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Bir bölgede yaşayan ailelerin yıllık ortalama gelirleri ile ulaşım harcamaları aşağıda verilmiştir. (1=1000 TL)

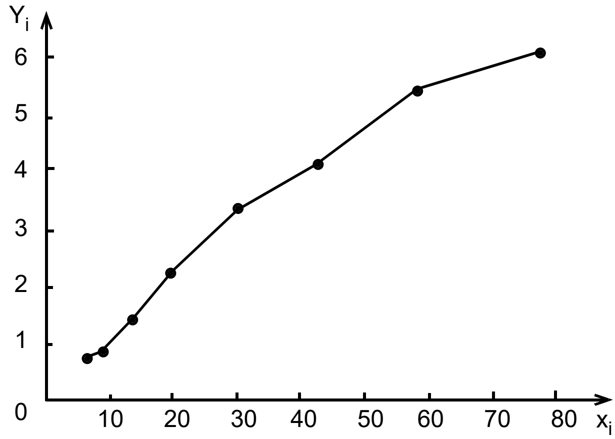
Yıllar	Gelir	Ulaşım H.
2002	5,2	0,7
2003	8,5	0,9
2004	13,4	1,5
2005	20,2	2,3
2006	30,5	3,4
2007	42,0	4,1
2008	59,3	5,2
2009	80,9	5,9
	260,0	24,0

a) Serinin grafiğini çiziniz.
b) Basit doğrusal regresyon modelinin parametrelerini tahmin ediniz.
c) Parametre tahmincilerinin varyanslarını hesaplayınız.
d) Korelasyon katsayısını hesaplayınız.
e) Belirlilik katsayısını hesaplayınız.
f) 2010 yılında ailelerin ortalama gelirlerinin 2009 yılına göre % 60 artacağı varsayımı ile 2010 yılı ortalama ulaşım harcamalarını tahmin ederek, tahminin standart hatasını hesaplayınız.
g) Sabit ve eğim katsayılarını 0,05 hata payı ile çift taraflı olarak test edin.
h) F testini yaparak t_{β}^{2} =F bağıntısını gösterin.

Çözüm:

Y_i	X_i	X_iY_i	X_i^2	Y_i^2
0,7	5,2	3,64	27,04	0,49
0,9	8,5	7,65	72,25	0,81
1,5	13,4	20,10	179,56	2,25
2,3	20,2	46,46	408,04	5,29
3,4	30,5	103,70	930,25	11,56
4,1	42,0	172,20	1764,00	16,81
5,2	59,3	308,26	3516,49	27,04
5,9	80,9	477,31	6544,81	34,81
24,0	260,0	1139,42	13442,44	99,06

a)



b) $\overline{X} = \dfrac{\sum X_i}{n} = \dfrac{260}{8} = 32{,}5$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{24}{8} = 3$$

$$\hat{\beta} = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

$$= \frac{1139{,}42 - 8(32{,}5)(3)}{13442{,}44 - 8(32{,}5)^2}$$

$$= \frac{359{,}42}{4992{,}44} = 0{,}071$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= 3 - 0{,}071(32{,}5) = 0{,}692$$

$$Y_i = 0{,}692 + 0{,}071 X_i$$

Gelirin bin Türk lirası artması halinde ulaşım harcamaları (0,071 x 1.000 =) 71 TL artmaktadır. Gelir olmasa bile ($X_i = 0$) (0,692 x 1.000 =) 692 TL ulaşım harcaması yapılacaktır.

a) $S_e^2 = \dfrac{\sum e_i^2}{n-2}$

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$$

$$= 99{,}06 - 0{,}692(24) - 0{,}071(1139{,}42)$$

$$=1{,}553$$

$$S_e^2 = \frac{1,553}{8-2} = 0,258$$

$$S_{\hat{\alpha}}^2 = S_e^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right]$$

$$\sum (X_i - \overline{X})^2 = \sum X_i^2 - n\overline{X}^2$$

$$= 13442,44 - 8(32,5)^2 = 4992,44$$

$$S_{\hat{\alpha}}^2 = 0,258 \left[\frac{1}{8} + \frac{(32,5)^2}{4992,44} \right]$$

$$= 0,08683$$

$$S_{\hat{\beta}}^2 = \frac{S_e^2}{\sum (X_i - \overline{X})^2}$$

$$= \frac{0,258}{4992,44} = 0,000051$$

d)

$$r = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\sqrt{[\sum X_i^2 - n\overline{X}^2][\sum Y_i^2 - n\overline{Y}^2]}}$$

$$= \frac{1139,42 - 8(32,5)(3)}{\sqrt{[13442,44 - 8(32,5)^2][99,06 - 8(3)^2]}}$$

$$= \frac{359,42}{\sqrt{(4992,44).(27,06)}}$$

$$= \frac{359,42}{367,55} = 0,977$$

e) $R^2 = r^2 = (0,977)^2 = 0,954$

ve

$$\sum e_i^2 = 1,553$$

$$\sum (Y - \overline{Y})^2 = \sum Y_i^2 - n\overline{Y}^2$$

$$= 99,06 - 8(3)^2 = 27,06$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \overline{Y})^2}$$

$$= 1 - \frac{1{,}553}{27{,}06}$$

$$= 1 - 0{,}057 = 0{,}943$$

Gelir, ulaşım harcamalarındaki değişmelerin yaklaşık % 95'ini açıklamaktadır.

f) Artış oranı 0,60 olursa 2010 yılı geliri,

(80,9)(1,6) = 129,44

olacaktır.

$$\hat{Y}_{10} = 0{,}692 + 0{,}071\ X_{10}$$

$$= 0{,}692 + 0{,}071\ (129{,}44) = 9{,}882$$

$$S_T = \sqrt{S_e^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum (X_i - \overline{X})^2} \right]}$$

$$= \sqrt{0{,}258 \left[1 + \frac{1}{8} + \frac{(129{,}44 - 32{,}5)^2}{4992{,}44} \right]}$$

$$= \sqrt{0{,}77588} = 0{,}880$$

g) $Y_i = 0{,}692 + 0{,}071 X_i$

$S_{\hat{\alpha}}^2 = 0{,}08683$ $\qquad$ $S_{\hat{\alpha}} = 0{,}2946$

$S_{\hat{\beta}}^2 = 0{,}000051$ $\qquad$ $S_{\hat{\beta}} = 0{,}0071$

olarak bulunmuştur.

α 'nın testi:

$H_0 : \alpha = 0$

$H_1 : \alpha \neq 0$

$t_{\alpha/2,n-2} = t_{0,05/2,8-2} = 2{,}447$

$$t = \frac{\hat{\alpha}}{S_{\hat{\alpha}}} = \frac{0{,}692}{0{,}2946} = 2{,}348$$

$|2{,}348| < |2{,}447|$ olduğundan H_0 hipotezi kabul edilir.

β 'nın testi:

$H_0: \beta = 0$

$H_1: \beta \neq 0$

$t_{\alpha/2,n-2} = t_{0,05/2,8-2} = 2,447$

$$t = \frac{\hat{\beta}}{S_{\hat{\beta}}} = \frac{0,071}{0,0071} = 10$$

$|10| > |2,447|$ olduğundan H_1 hipotezi kabul edilir.

h) $H_0: \beta = 0$

$H_1: \beta \neq 0$

$F_{\alpha,1,n-2} = F_{0,05,1,6} = 5,99$

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-2}{1}$$

$$= \frac{0,943}{1-0,943} \cdot \frac{6}{1} = 99,26$$

$|99,26| > |5,99|$ olduğundan H_1 hipotezi kabul edilir.

$t_{\beta} = 10$

$t_{\beta}^{2} = (10)^2 = 100 \cong 99,26$

sapma işlemlerden kaynaklanmaktadır.

ÖRNEK 2: Bir ailenin yedi yıllık geliri ve gıda harcamaları aşağıda verilmiştir. (1 = 100 T.L.)

Yıllar	Gelir	Gıda Harcamaları
2003	13	3
2004	18	4
2005	24	6
2006	31	9
2007	25	8
2008	45	15
2009	54	11
	210	56

a) $Y_i = \beta_0 + \beta_1 X_i + e_i$ regresyon modelinin parametrelerini tahmin ediniz.

b) Hata terimlerinin varyansını tahmin ediniz.

c) Korelasyon katsayısını hesaplayınız.

d) Belirlilik katsayısını hesaplayınız.

e) Ailenin 2010 yılı gelirinin en az 650 T.L., ve en çok 800 T.L. olacağını tahmin edildiğine göre kişinin 2010 yılı gıda harcamalarını tahmin ediniz.

Çözüm:

X_i	Y_i	$X_i - \overline{X}$	$Y_i - \overline{Y}$	$(X_i - \overline{X})(Y_i - \overline{Y})$	$(X_i - \overline{X})^2$	$(Y_i - \overline{Y})^2$
13	3	-17	-5	85	289	25
18	4	-12	-4	48	144	16
24	6	-6	-2	12	36	4
31	9	1	1	1	1	1
25	8	-5	0	0	25	0
45	15	15	7	105	225	49
54	11	24	3	72	576	9
210	56	0	0	323	1296	104

a) $$\overline{X} = \frac{\sum X_i}{n} = \frac{210}{7} = 30$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{56}{7} = 8$$

$$\hat{\beta} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

$$= \frac{323}{1296} = 0,249$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

= 8 - 0,249 (30) = 0,53

$Y_i = 0,53 + 0,249\ X_i$

Gelirin 100 TL artması gıda harlamalarını 100 x 0,249 = 24,9 lira arttıracaktır. Ailenin gelirinin olmaması durumunda gıda harcamaları 100 x 0,53 = 53 TL olacaktır.

b) $$S_e^2 = \frac{\sum e_i^2}{n-2}$$

$$\sum e_i^2 = \sum (Y_i - \overline{Y})^2 - \hat{\beta} \sum (X_i - \overline{X})(Y_i - \overline{Y})$$

= 104 –0,249(323) =23,573

$$S_e^2 = \frac{23,573}{7-2} = 4,7146$$

c) $r = \dfrac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}$

$= \dfrac{323}{\sqrt{1296.104}}$

$= \dfrac{323}{367,12} = 0,879$

Gelir ile gıda harcamaları arasında doğru yönlü kuvvetli ilişki vardır.

d) $R^2 = r^2 = (0,879)^2 = 0,772$

veya

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y - \overline{Y})^2}$$

$$= 1 - \frac{23,573}{104}$$

$= 1 - 0,226 = 0,774$

Gelir, gıda harcamalarındaki değişimin %77,4 ünü açıklamaktadır.

e) $Yi = 0,53 + 0,24\,9\,X_i$

$\hat{X}_{10,1} = 650$

$\hat{X}_{10,2} = 800$

$\hat{Y}_{10,1} = 0,53 + 0,249(650) = 167,15$

$\hat{Y}_{10,2} = 0,53 + 0,249(800) = 204,5$

Ailenin 2010 yılı geliri 650 TL olursa gıda harcamaları 100 x 167,15 = 16715 lira, 800 TL olursa gıda harcamaları 100 x 204,5 = 20450 TL olacaktır.

ÖRNEK 3: Aşağıda bir işletmenin 12 yıllık satışları ve karları ile ilgili olarak aşağıdaki değerler hesaplanmıştır. (1=100 T.L.)

$$\sum_{i=1}^{12} Y_i = 96 \qquad \sum_{i=1}^{12} Y_i^2 = 1064 \qquad \sum_{i=1}^{12} X_i Y_i = 6564$$

$$\sum_{i=1}^{12} X_i = 666 \qquad \sum_{i=1}^{12} X_i^2 = 42922$$

Bu değerlerden yararlanarak,

a) Basit doğrusal regresyon modelinin parametrelerini tahmin ediniz.

b) Korelasyon katsayısını hesaplayınız.

c) Belirlilik katsayısını hesaplayınız.

Çözüm:

a) $\overline{Y} = \dfrac{\sum Y_i}{n} = \dfrac{96}{12} = 8$

$$\overline{X} = \frac{\sum X_i}{n} = \frac{666}{12} = 55{,}5$$

$$\hat{\beta} = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

$$= \frac{6564 - 12(55{,}5)(8)}{42922 - 12(55{,}5)^2}$$

$$= \frac{1236}{5959} = 0{,}207$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

= 8 – 0,207(55,5) = -3,488

Y_i = - 3,488 + 0,207 X_i

Satışların 100 T.L. artışı, karı (100)(0,207) = 20,7 T.L. arttıracaktır. Satış yapılmadığında işletmenin zararı (100) (-3,488) = - 348,8 - T.L. olacaktır.

b) $r = \dfrac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\sqrt{[\sum X_i^2 - n(\overline{X})^2][\sum Y_i^2 - n(\overline{Y})^2]}}$

$$= \frac{6564 - 12(55,5)(8)}{\sqrt{[42922 - 12(55,5)^2][1064 - 12(8)^2]}}$$

$$= \frac{1236}{\sqrt{5959.296}} = 0{,}930$$

satışlar ve kar arasında doğru yönlü kuvvetli ilişki vardır.

c) $R^2 = r^2 = (0{,}93)^2 = 0{,}8649$

veya

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$$

= 1064 - (-3,488) (96) - (0,207) (6564) = 40,1

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \overline{Y})^2}$$

$$\sum(Y_i-\overline{Y})^2=\sum Y_i^2-n(\overline{Y})^2$$

$$= 1064 - 12(8)^2 = 296$$

$$R^2=1-\frac{40{,}1}{296}$$

$$= 1 - 0{,}135 = 0{,}865$$

Satışlar kardaki değişmelerin % 86, 5'ini açıklamaktadır.

ÖRNEK 4: Aşağıda bir malın üretim miktarları ile maliyetleri verilmiştir.

Maliyet (1=1000 TL)	Üretim Miktarı (1=1000 Adet)
22	1
18	2
16	3
14	4
12	7
11	10

a) $Y_i=\alpha+\beta X_i$

b) $Y_i=\beta_0+\beta_1 X_i+\beta_2 X_i^2$

c) $Y_i=\alpha+\beta\log X_i$

d) $\log Y_i=\alpha+\beta X_i$

e) $\log Y_i=\alpha+\beta\log X_i$

modellerini tahmin ederek, belirlilik katsayılarını hesaplayınız.

Çözüm:

X_i	Y_i	X_iY_i	X_i^2	Y_i^2	X_i^3	X_i^4	$X_i^2Y_i$
1	22	22	1	484	1	1	22
2	18	36	4	324	8	16	72
3	16	48	9	256	27	81	144
4	14	56	16	196	64	256	224
7	12	84	49	144	343	2401	588
10	11	110	100	121	1000	10000	1100
27	93	356	179	1525	1443	12755	2150

a) $Y_i=\hat{\alpha}+\hat{\beta}X_i$

$$\overline{X}=\frac{\sum X_i}{n}=\frac{27}{6}=4{,}5$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{93}{6} = 15{,}5$$

$$\hat{\beta} = \frac{\sum X_i Y_i - n\overline{XY}}{\sum X_i^2 - n(\overline{X})^2}$$

$$= \frac{356 - 6(4{,}5)(15{,}5)}{179 - 6(4{,}5)^2}$$

$$= \frac{-62{,}5}{57{,}5} = -1{,}086$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

= 15,5 – (-1,086)(4,5) = 20,387

Y_i = 20,387 – 1,086 X_i

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$$

= 1525 – 20,387(93) – (1,086)(356) =15,625

$$\sum (Y_i - \overline{Y})^2 = \sum Y_i^2 - n(\overline{Y})^2$$

= 1525 – 6(15,5)^2 =83,5

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \overline{Y})^2}$$

$$R^2 = 1 - \frac{15{,}625}{83{,}5}$$

= 1 – 0,187 = 0,813

b) $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2$

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

$$\sum X_i Y_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 + \hat{\beta}_2 \sum X_i^3$$

$$\sum X_i^2 Y_i = \hat{\beta}_0 \sum X_i^2 + \hat{\beta}_1 \sum X_i^3 + \hat{\beta}_2 \sum X_i^4$$

$$-4{,}5/93 = 6\hat{\beta}_0 + 27\hat{\beta}_1 + 179\hat{\beta}_2$$

$$356 = 27\hat{\beta}_0 + 179\hat{\beta}_1 + 1443\hat{\beta}_2$$

$$-418{,}5 = -27\hat{\beta}_0 - 121{,}5\hat{\beta}_1 - 805{,}5\hat{\beta}_2$$

$$356 = 27\hat{\beta}_0 + 179\hat{\beta}_1 + 1443\hat{\beta}_2$$

$$-62{,}5 = 0 + 57{,}5\hat{\beta}_1 + 637{,}5\hat{\beta}_2$$

$$-179/93 = 6\hat{\beta}_0 + 27\hat{\beta}_1 + 179\hat{\beta}_2$$

$$6/2150 = 179\hat{\beta}_0 + 1443\hat{\beta}_1 + 12755\hat{\beta}_2$$

$$-16647 = -1074\hat{\beta}_0 - 4833\hat{\beta}_1 - 32041\hat{\beta}_2$$

$$12900 = 1074\hat{\beta}_0 + 8658\hat{\beta}_1 + 76530\hat{\beta}_2$$

$$-3747 = 0 + 3825\hat{\beta}_1 + 44489\hat{\beta}_2$$

$$-3825/-62{,}5 = 57{,}5\hat{\beta}_1 + 637{,}5\hat{\beta}_2$$

$$57{,}5/-3747 = 3825\hat{\beta}_1 + 44489\hat{\beta}_2$$

$$239062{,}5 = -219937{,}5\hat{\beta}_1 - 2438437{,}5\hat{\beta}_2$$

$$-215452{,}5 = 219937{,}5\hat{\beta}_1 + 2558117{,}5\hat{\beta}_2$$

$$23610 = 119680\hat{\beta}_2$$

$$\hat{\beta}_2 = \frac{23610}{119680} = 0{,}197$$

$$-62{,}5 = 57{,}5\hat{\beta}_1 + 637{,}5(0{,}197)$$

$$\hat{\beta}_1 = \frac{-62{,}5 - 125{,}5875}{57{,}5} = -3{,}271$$

$$93 = 6\,\hat{\beta}_0 + 27(-3{,}271) + 179(0{,}197)$$

$$\hat{\beta}_0 = \frac{93 + 88{,}317 - 35{,}263}{6} = 24{,}342$$

$$Y_i = 24{,}342 - 3{,}217X_i + 0{,}197X_i^2$$

$$\sum e_i^2 = \sum Y^2 - \hat{\beta}_0 \sum Y_i - \hat{\beta}_1 \sum X_i Y_i - \hat{\beta}_2 \sum X_i^2 Y_i$$

= 1525 –24,342(93) – (-3,271)(356) – 0,197(2150)

=2,12

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

$$= 1 - \frac{2,12}{83,5}$$ = 1 - 0,025 =0,975

c) $Y_i = \hat{\alpha} + \hat{\beta} \log X_i$

$$\hat{\beta} = \frac{\sum Y_i (\log X_i) - n\bar{Y}(\overline{\log X})}{\sum (\log X)^2 - n(\overline{\log X})^2}$$

$\log X_i$	$(\log X_i)^2$	$Y_i(\log X_i)$
0,00000	0,00000	0,00000
0,30103	0,09061	5,41854
0,47712	0,22764	7,63392
0,60205	0,36246	8,42870
0,84509	0,71418	10,14108
1,00000	1,00000	11,00000
3,22529	2,39489	42,62224

$$(\overline{\log X}) = \frac{\sum \log X_i}{n} = \frac{3,22529}{6} = 0,53754$$

$$\beta = \frac{42,62224 - 6(15,5)(0,53754)}{2,39489 - 6(0,53754)^2}$$

$$= \frac{-7,36898}{0,66119} = -11,14503$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}(\overline{\log X})$$

= 15,5 – (-11,14503)(0,53754) = 21,49089

Y_i = 21,49089 – 11,14503 $\log X_i$

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha} \sum Y_i - \hat{\beta} \sum Y_i (\log X_i)$$

= 1525 – 21,49089(93) – (-11,14503)(42,62224)

= 1,37337

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \overline{Y})^2}$$

$$= 1 - \frac{1{,}37337}{83{,}5}$$

$$= 1 - 0{,}016 = 0{,}984$$

d) $\log X_i = \hat{\alpha} + \hat{\beta} X_i$

$$\beta = \frac{\sum X_i(\log Y_i) - n\overline{X}(\overline{\log Y})}{\sum X^2 - n(\overline{X})^2}$$

$\log Y_i$	$(\log Y_i)^2$	$X_i(\log Y_i)$
1,34242	1,80209	1,34242
1,25527	1,57570	2,51054
1,20412	1,44990	3,61236
1,14612	1,31360	4,58448
1,07918	1,16463	7,55426
1,04139	1,08449	10,41390
7,06850	8,39041	30,01796

$$(\overline{\log Y}) = \frac{\sum \log Y_i}{n} = \frac{7{,}06850}{6} = 1{,}17808$$

$$\hat{\beta} = \frac{30{,}01796 - 6(4{,}5)(1{,}17808)}{179 - 6(4{,}5)^2} = \frac{-1{,}7902}{57{,}5} = -0{,}03113$$

$$\hat{\alpha} = (\overline{\log Y}) - \hat{\beta}\overline{X}$$

$$= 1{,}17808 - (-0{,}03113)(4{,}5) = 1{,}31816$$

$\log Y_i = 1{,}31816 - 0{,}03113 X_i$

$$\sum e_i^2 = \sum (\log Y_i)^2 - \hat{\alpha}\sum \log Y_i - \hat{\beta}\sum X_i(\log Y_i)$$

$$= 8{,}39041 - 1{,}31816(7{,}06850) - (-0{,}03113)(30{,}01796)$$

$$= 0{,}00745$$

$$\sum [\log Y_i - (\overline{\log Y})]^2 = \sum (\log Y_i)^2 - n(\overline{\log Y})^2$$

$$= 8{,}39041 - 6(1{,}17808)^2 = 0{,}06317$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum [\log Y - (\overline{\log Y})]^2} = 1 - \frac{0{,}00745}{0{,}06317}$$

$$= 1 - 0{,}117 = 0{,}883$$

e) $\log X_i = \log \hat{\alpha} + \hat{\beta} \log X_i$

$$\beta = \frac{\sum (\log X_i)(\log Y_i) - n(\overline{\log X})(\overline{\log Y})}{\sum (\log X)^2 - n(\overline{\log X})^2}$$

$(\log X_i)(\log Y_i)$
0,00000
0,37787
0,57450
0,60002
0,91200
1,04139
3,59578

$$\hat{\beta} = \frac{3{,}59578 - 6(0{,}53754)(1{,}17808)}{2{,}39489 - 6(0{,}53754)^2}$$

$$= \frac{-0{,}20381}{0{,}66119} = -0{,}30824$$

$$\log \hat{\alpha} = (\overline{\log Y}) - \hat{\beta}(\overline{\log X})$$

$= 1{,}17808 - (-0{,}30824)(0{,}53754) = 1{,}34377$

$\log Y_i = 1{,}34377 - 0{,}30824(\log X_i)$

$$\sum e_i^2 = \sum (\log Y_i)^2 - (\log \hat{\alpha}) \sum \log Y_i - \hat{\beta} \sum (\log X_i)(\log Y_i)$$

$= 8{,}39041 - 1{,}34377(7{,}06850) - (-0{,}30824)(3{,}59578)$

$= 0{,}000335$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum [\log Y_i - (\overline{\log Y})]^2}$$

$$= 1 - \frac{0{,}000335}{0{,}06317}$$

$= 1 - 0{,}0053 = 0{,}9947$