

KORELASYON ve REGRESYON

Birçok durumda araştırmacı üzerinde çalıştığı örnekte birden fazla özelliğe ait veri toplayabilir. Bu gibi durumlarda sadece örneğin tanıtıcı istatistiklerini hesaplamak ve bilmek araştırmacı için yeterli olmayacaktır. Çünkü bu elde veriler olmasına rağmen bilgi kaybı demektir ve eldeki veriler kullanılarak iki özellik arasında bir ilişki olup olmadığı, değişkenlerden birinin bir birim artmasına karşılık diğer değişkende nasıl bir değişiklik meydana geldiği araştırılabilir.

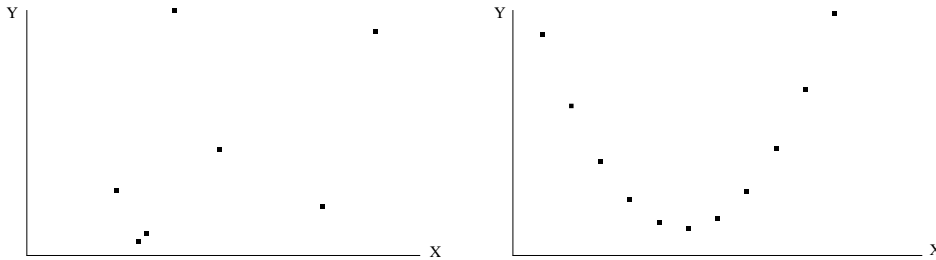
Bir örnekten birden fazla özelliğe ait veri toplandığı zaman hesaplanması gereken istatistikler korelasyon ve regresyon katsayılarıdır. Bu kitapta söz konusu katsayılar açıklanırken iki özellik arasındaki doğrusal ilişki dikkate alınmıştır.

Değişkenler arasındaki ilişkilerin kaynağı değişik olabilir. Ele alınan değişkenlerden biri diğerini etkileyen etkenlerden (faktörlerden) biri olabilir. Bu tip ilişkilere **sebe-sonuç** ilişkisi denir. Ele alınan iki değişkenin her ikisini de etkileyen bir veya birçok faktörün varlığı da ilişki sebebidir. Her ilişkiyi sebe-sonuç ilişkisi olarak ele almak yanlıştır¹. Örneğin, sigara tüketimi ile akciğer kanseri vakası sayısı arasındaki ilişkiyi araştırırken, 20 yıl boyunca tüketilen sigara miktarı ile akciğer kanserinden ölenlerin sayısı arasında hesaplanan ilişkiyi dikkatli yorumlamak gerekir. Bu ilişkide sigaranın kanser yapma etkisi yanında, zamana bağlı olarak nüfusun, teşhis metotlarındaki gelişmenin ve doktora başvurma alışkanlığının artmasının da etkisi vardır.

5.1. Korelasyon Katsayısı

Bir örnekten iki özelliğe ait toplanan verilerin koordinat sisteminde noktalar halinde gösterilmesi araştırmacıya iki özellik arasında bir ilişki olup olmadığı, eğer varsa ilişkinin şekli hakkında bir ön bilgi verecektir.

Eğer veriler koordinat ekseninde işaretlendiği zaman Şekil 5.1’de verilen grafiklerde görüldüğü gibi bir dağılım gösteriyorsa iki özellik arasındaki ilişkinin derecesini belirten korelasyon katsayısı (r) 0 veya 0’a çok yakın bir değerdir. Bu iki özellik arasında doğrusal bir ilişki olmadığını gösterir.



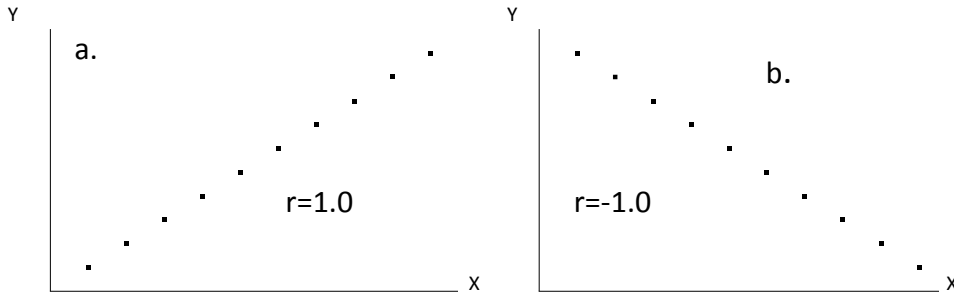
ŞEKİL 5.1. X ve Y özellikleri arasındaki ilişkinin 0 veya 0’a yakın olabileceği durumlar

¹ Birçok değişkenin zaman periyodu içinde değişmesi de bu değişkenler arasında bir ilişkiye neden olur.

4. HAFTA DERS NOTLARI

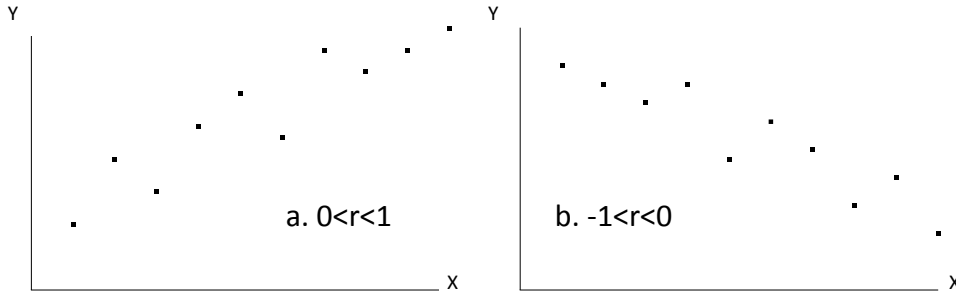
Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

İki özelliğe ait veriler koordinat sisteminde işaretlendiği zaman aşağıdaki grafiklerde gösterildiği gibi noktalar bir doğru üzerinde sıralanabilir. Bu, iki özellik arasında tam bir ilişki olduğunu gösterir, yani her X değerine karşılık gelen bir tek Y değeri vardır. Tam ilişki olduğu zaman korelasyon katsayısı 1.0 olarak bulunur. İlişkinin yönünü ise korelasyon katsayısının işareti belirtir. Eğer iki özellik arasındaki ilişki Şekil 5.2.a'daki gibi artan bir ilişki ise yani X arttıkça Y de artıyorsa korelasyon pozitif işaretlidir, yani $r=1.0$ olarak bulunur. Fakat iki özellik arasındaki ilişki Şekil 5.2.b'de verildiği gibi negatif (ters, azalan) bir ilişki ise korelasyon katsayısı negatif işaretlidir, yani $r=-1.0$ olarak bulunur.



ŞEKİL 5.2. X ve Y özellikleri arasında a. artan ve b. azalan tam doğrusal ilişki

Veriler koordinat sisteminde işaretlendiği zaman Şekil 5.3.a ve b'de verildiği gibi de olabilir.



ŞEKİL 5.3. X ve Y özellikleri arasında tam olmayan a. artan ve b. azalan doğrusal ilişki

Şekil 5.3'de görüldüğü gibi iki özellik arasında tam bir ilişki yoktur, yani noktalar bir doğru üzerinde dizilmemişlerdir. Bu durumda, eğer Şekil 5.3.a'da olduğu gibi iki özellik arasında tam olmayan pozitif bir ilişki varsa korelasyon katsayısı 0 ile +1 arasında bir değer alır. İki özellik arasında negatif tam olmayan bir ilişki söz konusu olduğu zaman ise (Şekil 5.3.b) korelasyon katsayısı -1 ile 0 arasındadır.

Yukarıda da açıklandığı gibi korelasyon katsayısı -1 ile +1 arasında değişir. İlişki tam ise 1 değerini alır. Korelasyon katsayısının işareti ise ilişkinin yönünü gösterir. Korelasyon katsayısının eşiti ise popülasyonda aşağıdaki gibidir:

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Populasyonda :

$$\rho = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum (x_i - \mu_x)^2 \sum (y_i - \mu_y)^2}} \quad \dots(5.1)$$

Örnekte

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad \dots(5.2)$$

Ve kısaca aşağıdaki gibi gösterilir.

$$r = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}} \quad \dots(5.3)$$

Eşitlikte, $\sum d_x d_y$, çarpımlar toplamı, $\sum d_x^2$, X değişkenine ait kareler toplamı, $\sum d_y^2$, Y değişkenine ait kareler toplamıdır.

Çarpımlar toplamı, X değerlerinin kendi ortalamalarından sapmaları ile Y değerlerinin kendi ortalamalarından sapmalarının çarpımlarının toplamıdır ve $\sum d_x d_y = \sum (x_i - \bar{x})(y_i - \bar{y})$ şeklinde gösterilir. Kareler toplamının hesaplanmasında olduğu gibi çarpımlar toplamının hesaplanmasında da kolaylık için aşağıda verilen eşitlik kullanılır. Her birey için X ve Y değerlerinin çarpımlarının toplamından X ve Y değerlerinin toplamlarının çarpımının birey sayısına bölümü çıkarılır. Korelasyon katsayısının işaretini, çarpımlar toplamının işareti belirler.

$$\sum d_x d_y = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} \quad \dots(5.4)$$

5.2. Regresyon Katsayısı

Ele alınan iki değişkenden biri diğerinin fonksiyonu olarak ele alınabilir. Bu durum özelliklerin doğası olarak böyle olabildiği gibi, araştırmacı işine uygun olduğu için kendisi böyle seçmiş olabilir. Bir ülkenin nüfusu yıllar ilerledikçe artar. Yani pozitif bir ilişki vardır. Burada ülke nüfusu yılların (zamanın) bir fonksiyonudur. Nüfus Y, zaman X ile tanımlanırsa bu ilişki kısaca **y=f(x)** olarak gösterilir. Herhangi bir aracın fren sonucu yerde bıraktığı iz (metre olarak) o andaki hızının (km/saat) bir fonksiyonudur. Gerçekte ilişki **fren mesafesi = f(hız)** şeklindedir. Ancak günlük yaşamda herhangi bir kaza sonucu aracın o andaki hızı bilinmek istenir. Bu kaza yapan taraflar için çok önemlidir. Ancak kaza olmuş bitmiştir ve kanıt olarak yolda aracın bıraktığı fren izi vardır. İşte hızı tahmin etmek için gerçekteki durumun aksine **hız=f(fren mesafesi)** şeklinde ele alınabilir.

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Değişkenlerden biri diğerinin fonksiyonu olarak tanımlandığında eşitliğin sol tarafındaki değişkene **bağımlı** değişken sağ tarafındakine de **bağımsız** değişken denir. Bağımlı değişken birden fazla değişkenin fonksiyonu olarak da ele alınabilir, $y=f(x_1, x_2, x_3)$ gibi.

Bu bölümde bir bağımlı ve bir bağımsız değişken arasındaki ilişki ele alınacaktır.

$y=f(x)$ eşitliğinde; X bağımsız değişken Y ise bağımlı değişkendir. Örnekteki X ve Y çiftleri koordinat ekseninde noktalandığında, ilişkinin doğrusal olduğunun varsayılabileceği kararına varılmış ise fonksiyon örnekte;

$$Y=a+bX+e \text{ şeklinde ifade edilebilir.}$$

Bağımsız değişkenin kendi ölçü birimi cinsinden bir birim değişmesine karşılık bağımlı değişkenin kendi ölçü birimi cinsinden ortalama olarak ne kadar değişeceğini gösteren katsayıya regresyon katsayısı denir. Korelasyon katsayısı iki özellik arasında ilişkinin derecesini verir ve bir birimi yoktur. Fakat regresyon katsayısının birimi vardır ve bağımsız değişkende bir birim değişmeye karşılık bağımlı değişkenin kendi birimi cinsinden ortalama olarak değişeceği miktardır. Bağımsız değişken X ve bağımlı değişken Y ile gösterilirse: X değişkeninin bir birim artmasına karşılık Y değişkenin kendi birimi cinsinden ortalama olarak değişeceği miktara Y'nin X'e göre regresyon katsayısı denir ve b_{yx} olarak gösterilir. Eğer Y değişkeni bağımsız değişken ise bu durumda da b_{xy} şeklinde ifade edilir.

$$b_{yx}= Y'nin X'e göre regresyon katsayısı$$

$$b_{xy}= X'in Y'ye göre regresyon katsayısı$$

Regresyon katsayısının hesaplanması için kullanılan eşitlik genel biçimde aşağıdaki gibidir:

$$\text{Regresyon katsayısı (b)} = \frac{\text{Çarpımlar Toplamı}}{\text{Bağımsız değişkene ait kareler toplamı}}$$

Bu durumda b_{yx} katsayısının hesaplanması için kullanılacak eşitlik;

$$b_{yx} = \frac{\sum d_x d_y}{\sum d_x^2} \quad \dots(5.5)$$

ve b_{xy} katsayısının hesaplanması için ise kullanılacak eşitlik;

$$b_{xy} = \frac{\sum d_x d_y}{\sum d_y^2} \quad \dots(5.6)$$

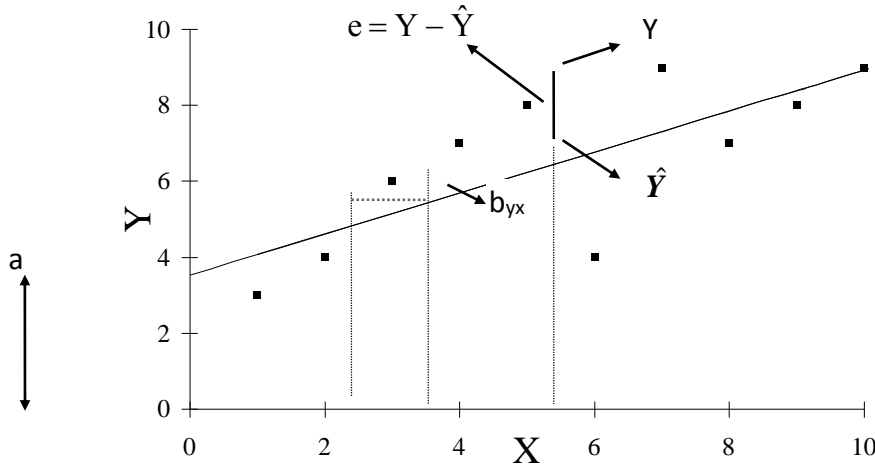
4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Yukarıda verilen eşitlikler kullanılarak regresyon katsayısı hesaplandığı zaman iki özellik arasındaki ilişkinin yönüne bağlı olarak katsayının işareti pozitif veya negatif çıkar. Eğer iki özellik arasında ters bir ilişki varsa katsayı negatif işaretli, artan bir ilişki varsa katsayı pozitif işaretli olarak bulunur. Regresyon katsayısının işareti ile korelasyon katsayısının işareti hiç bir zaman farklı bulunamaz çünkü her ikisinin de işareti çarpımlar toplamının işareti ile belirlenir.

5.3. Regresyon Denklemi

Eğer işaretlenen noktalar bir doğru üzerinde olsa idi doğrunun denklemi kolayca oluşturulabilirdi. Eğer noktalar tam bir doğru üzerinde değil de, doğru etrafında dağılıyorsa X ve Y değerlerine karşılık gelen noktaların hepsine birden en yakın geçecek bir doğru oluşturulabilir. Oluşturulacak bu doğruya “**Regresyon Doğrusu**”, bu doğrunun denklemine de “**Regresyon Denklemi**” veya “**Önceden Tahmin Denklemi**” denir ve $\hat{Y} = a + b_{yx}X$ şeklinde gösterilir. Şekil 5.4’te gösterildiği gibi denklemdeki “a” katsayısı, regresyon doğrusunun Y-eksenini kestiği noktanın ordinatı, b_{yx} ise regresyon doğrusunun eğimi yani regresyon katsayısıdır.



ŞEKİL 5.4. X ve Y özelliklerine ait gerçek gözlemler ve regresyon doğrusu

Regresyon denklemindeki katsayılar, X değerlerine karşılık gözlenen Y değerleri ile bu denklemden tahmin edilecek \hat{Y} değerleri arasındaki sapmaların kareler toplamını minimum yapacak şekilde hesaplanmışlardır, yani bu katsayılar $\sum(Y - \hat{Y})^2$ değerini minimum yapan değerlerdir. $\sum(Y - \hat{Y})^2$ eşitliğinde \hat{Y} yerine $(a + b_{yx}X)$ konarak $\sum(Y - (a + b_{yx}X))^2$ eşitliği elde edilir. Bu eşitliğin tahmin edilmek istenen “a” ve “b”ye göre kısmi türevleri alınarak sıfıra eşitlendiği ve elde edilen eşitlikler çözüldüğü zaman bu katsayıların eşitleri $a = \bar{Y} - b_{yx}\bar{X}$ ve $b_{yx} = \frac{\sum d_x d_y}{\sum d_x^2}$ olarak bulunur.

$$Y = a + b_{yx}X + e$$

ve;

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

tahmin eşitliği de $\hat{Y} = a + b_{yx}X$ olduğuna göre $Y - \hat{Y} = e$ 'dir. Buna regresyondan sapma denir. a ve b katsayıları, bu sapma kareler toplamını ($\sum e^2$) minimum yapan değerlerdir. Regresyondan sapmaların, ortalaması sıfır, varyansı σ^2 olan normal dağılım gösterdiği varsayılır.

5.4. İsbet (Doğruluk) Derecesi

Yukarıdaki tahmin edilen değerlerden de görüldüğü gibi gözlenen değerler ile tahmin edilen Y değerleri arasında bir farklılık vardır. Bunun sebebi iki özellik arasındaki ilişkinin tam ilişki olmamasıdır. Oluşturulan bu denklem ile yapılan tahminlerin doğruluk (isabet) derecesi, r^2 ile gösterilir. r^2 , yukarıda açıklanan ve hesaplanan korelasyon katsayısının karesine eşittir. Korelasyon katsayısının mutlak değer olarak 1'e yaklaşması regresyon denklemi kullanılarak yapılacak tahminlerin isabet derecesinin yüksek olduğu anlamına gelir.

ÖRNEK 1:

Herhangi bir vitamin karması şurubunda kapağı açıldıktan sonra oda sıcaklığında muhafazası halinde C vitamini miktarının değişmesini incelemek üzere yapılan bir çalışmada başlangıçta ve haftalar boyunca birim hacimde bulunan C vitamini miktarı aşağıdaki gibi olsun.

Haftalar	Birim hacimde bulunan C vitamini
0	100
2	90
4	70
6	40
8	30

Gözlemler koordinat ekseninde işaretlenince Şekil 5.5'deki grafik elde edilir.

Burada haftalar bağımsız değişken (X), buna bağlı olarak değişen vitamin miktarı ise bağımlı değişkendir (Y). Vitamin miktarı zamanın bir fonksiyonu olarak alınır, yani $Y=f(x)$ 'dir.

Şekil 5.5'te koordinat eksenine işaretlenen noktaların dağılımına bakılırsa, ilişkinin doğrusal olarak alınabileceği görülür. O halde varsayılan regresyon modeli;

$$Y = \alpha + \beta X + e$$

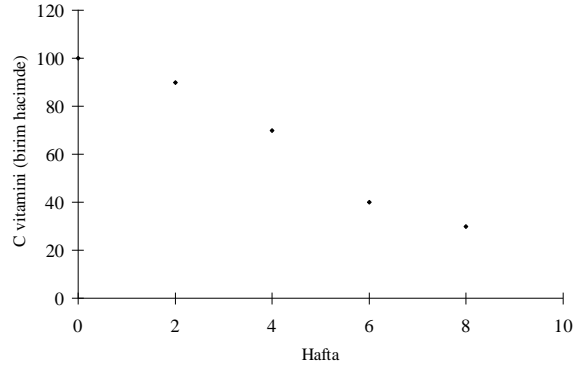
şeklindedir. Bu modelin ele alınan örnekten tahmini

$$\hat{Y} = a + b_{yx}X$$

şeklinde olacaktır.

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94



Şekil 5.5. Haftalar boyunca birim ünite de bulunan C vitamini miktarı

Bu modelin katsayılarının yukarıda açıklandığı şekilde hesaplanması için aşağıdaki tablo hazırlanmıştır.

X	Y	X ²	Y ²	XY
0	100	0	10000	0
2	90	4	8100	180
4	70	16	4900	280
6	40	36	1600	240
8	30	64	900	240
Toplam	20	330	120	25500

Toplamlar aşağıdaki gibidir:

$$\sum X = 20, \sum Y = 330, \sum XY = 940, \sum X^2 = 120, \sum Y^2 = 25500$$

$$\text{Bağımsız değişkenin ortalaması: } \bar{X} = \frac{20}{5} = 4$$

$$\text{Bağımlı değişkenin ortalaması: } \bar{Y} = \frac{330}{5} = 66$$

$$\text{Bağımsız değişkenin kareler toplamı: } \sum d_X^2 = 120 - \frac{(20)^2}{5} = 40$$

$$\text{Bağımlı değişkenin kareler toplamı: } \sum d_Y^2 = 25500 - \frac{(330)^2}{5} = 3720$$

$$\text{Çarpımlar toplamı: } \sum d_X d_Y = 940 - \frac{(20)(330)}{5} = -380$$

Birim hacimdeki C vitamini miktarının zamana göre regresyon katsayısı ise 5.5 numaralı eşitlikte yukarıda hesaplanan değerler yerine konarak aşağıdaki şekilde hesaplanır:

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

$$b_{yx} = \frac{-380}{40} = -9.5$$

-9.5 katsayısının anlamı “**Bağımsız değişkenin (haftanın) bir birim artmasına karşılık (1 hafta) bağımlı değişkenin (vitamin) 9.5 mg azaldığı**” şeklindedir.

$$a = \bar{Y} - b_{yx}\bar{X}$$

$$a = 66 - (-9.5)(4) = 104$$

Bu sonuçlara göre regresyon eşitliği,

$$\hat{Y} = 104 - 9.5X \text{ şeklindedir.}$$

Ele alınan zaman ve birim hacimdeki C vitamini arasındaki korelasyon katsayısı, 5.3 numaralı eşitlikte hesaplanan değerler yerine konarak aşağıdaki gibi hesaplanır:

$$r_{xy} = \frac{-380}{\sqrt{(40)(3720)}} = -0.985$$

Hesaplanan korelasyon katsayısı birim üitedeki C vitamini ile zaman arasında 0.985 gibi azalan doğrusal bir ilişki olduğunu gösterir. Bu örnek için belirtme katsayısı veya isabet derecesi de $(0.985)^2 \cong 0.97$ olarak bulunur. Bunun anlamı, bağımlı değişkendeki değişmelerin %97'sinin bağımsız değişken ile açıklanabildiği şeklindedir.

ÖRNEK 2:

Ağaçlardaki yaş halkalarının genişliği birçok faktörün etkisi altındadır. Bunlardan en önemlisi de yıllık yağış miktarıdır. Yağışın fazla olduğu yıllarda yaş halkaları daha geniş, az olduğu yıllarda ise daha dardır. Çok yaşlı ağaçların herhangi bir nedenle kesilme durumunda yaş halkalarının genişliklerinden yüzlerce yıl önceki yağış miktarları ve kurak geçen yıllar hakkında bilgi edinilebilir. Herhangi bir bölgede yağış gözlemi (rasadı) yapılan yıllar ve o bölgede kesilmiş olan ağacın o yıllara ait halka genişlikleri arasındaki ilişki belirlenerek 80–100 yıl öncesine (ağacın yaşına göre daha eski) ait yağış tahmin edilmeye çalışılır. Yağış halkası genişliğine (X), yıllık yağış miktarına da (Y) densin. Gerçekte yaş halkası sonuç, yıllık yağış miktarı da sebeptir. Yani **X=f(y)**'dir. Ancak yaş halkası genişliğinden yağış miktarı tahmin edilmek istendiğinden gerçek durumun aksine **yağış (mm)=f(halka genişliği)** veya **Y=f(x)** şeklinde, yani yağış, halka genişliğinin bir fonksiyonu şeklinde alınabilir.

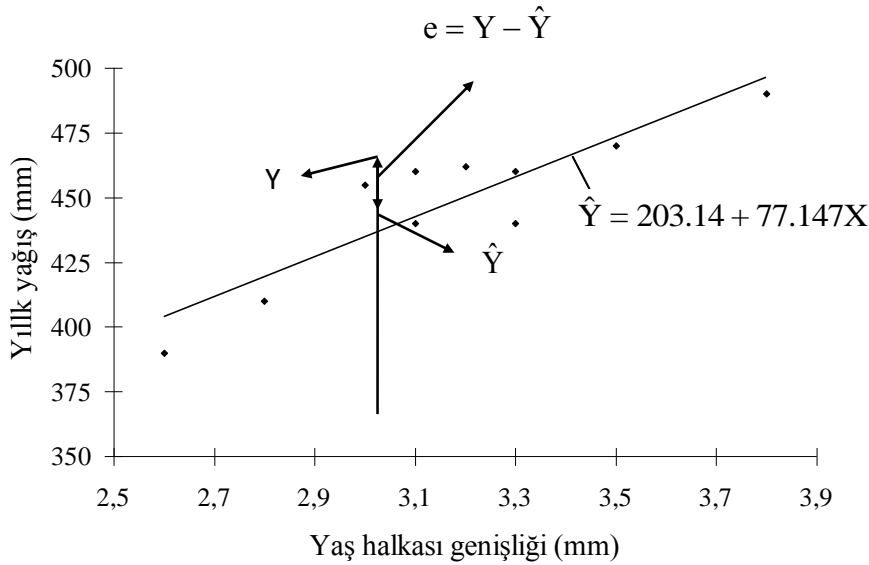
Örnek olarak herhangi bir bölgede yaşlı bir ağacın kesildiğini ve o bölgede son 10 yılda yıllık yağış rasadı (gerçekte daha uzun yılları kapsar) yapıldığını varsayalım. Bu yıllara ait yağış miktarları ve yaş halkası ölçümleri aşağıdaki gibi olsun. Yağış Y eksen, yağış halkası genişliği de X eksen olarak alındığında ve Şekil 5.6'da gösterildiği gibi gözlemler koordinat eksinine işaretlendiğinde $Y=a+bX$ gibi bir doğrusal modelin kurulabileceği görülmektedir.

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Yaş halkası genişliği (X)	Yıllık yağış (Y)	X ²	Y ²	XY
3.3	460	10.89	211600	1518.0
3.1	460	9.61	211600	1426.0
2.8	410	7.84	168100	1148.0
3.3	440	10.89	193600	1452.0
3.5	470	12.25	220900	1645.0
3.0	455	9.00	207025	1365.0
3.1	440	9.61	193600	1364.0
3.8	490	14.44	240100	1862.0
3.2	462	10.24	213444	1478.4
2.6	390	6.76	152100	1014.0
31.7	4477	101.53	2012069	14272.4

Not: Ağaçlar çok yaşlı olduğu için, bu ileri yaşlarda aynı yağış miktarında yağış halkası genişliğine yaştan etkisinin ihmal edilebilecek kadar az olduğu varsayılmıştır.



ŞEKİL 5.6. Yaş halkası genişliğine göre yıllık yağış miktarı

Gözlemlere ait aşağıdaki değerler hesaplanır:

Bağımsız değişkenin ortalaması: $\bar{X} = 3.17$

Bağımlı değişkenin ortalaması: $\bar{Y} = 447.7$

Bağımsız değişkenin kareler toplamı: $\sum d_X^2 = 1.041$

Bağımlı değişkenin kareler toplamı: $\sum d_Y^2 = 7716.1$

Çarpımlar toplamı: $\sum d_X d_Y = 80.31$

$\hat{Y} = a + b_{yx}X$ tahmin eşitliğinin katsayıları aşağıdaki gibi bulunmuştur:

$$b_{yx}=77.15 \quad a=203.14$$

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Bu örnek için hesaplanan regresyon denklemi ve iki özellik arasındaki korelasyon katsayısı ise aşağıdaki gibidir:

$$\hat{Y} = 203.14 + 77.15X$$

$$r = 0.90$$

Burada 77.15'in anlamı, yaş halkalarının 1 mm artmasına karşılık yağışta ortalama olarak 77.15 mm artmaktadır. Tahmin doğrusunu koordinat ekseninde çizmek için bağımsız değişkenin iki değerine karşılık bağımlı değişkenin aldığı değerler hesaplanır, bu noktalar doğru ile birleştirilir. Örnek olarak,

$$X = 2.7 \quad \text{için} \quad \hat{Y} = 411.45$$

$$X = 3.8 \quad \text{için} \quad \hat{Y} = 496.31$$

Doğruyu daha duyarlı çizmek için tahmin noktalarını çok yakın almamak gerekir. Çizilen bu doğru gözlenen noktaların hepsine birden en yakın geçen doğrudur (Şekil 5.6).

Koordinat eksenine işaretlenen noktalardan X eksenine dik indirildiğinde nokta ile doğru arasındaki uzaklığa regresyondan sapma denir, e_i gösterilir (Şekil 5.6). a ve b katsayıları bu sapmaların karelerinin toplamı en küçük olacak şekilde (en küçük kareler yöntemi) hesaplandığı daha önce açıklanmıştı. Yani, $\sum e_i^2$ minimumdur.

Katsayıları belirlenmiş olan tahmin eşitliğinden geçmiş yılların yağış miktarları tahmin edilebilir. 70 yıl önceki bir yılın yaş halkası genişliği 3.6 mm ise, o yılki tahmini yağış miktarı $\hat{Y} = 203.14 + 77.15(3.6) = 480.88$ mm'dir. Bu tahmindeki isabet derecesi korelasyon katsayısının karesine eşittir. Ele alınan örnekte belirtme katsayısı (isabet derecesi), $r^2=0.81$ 'dir.

Burada dikkate edilmesi gereken nokta, yaş halkaları genişliğinden yapılacak tahminde, tahmin denkleminde bağımsız değişken yerine konacak değer gözlem aralığında olması gereğidir. Ele alınan örnekte yaş halkası genişliği 2.6–3.8 mm arasında olmalıdır. Ancak bu değerlerden çok uzakta olmayan noktalar için de tahmin yapılabilir.

ÖRNEK 3:

Bundan önceki örnekte gerçek durumda $X=f(y)$ olduğu halde, pratik uygulamada daha kullanışlı olduğu için tahmin modeli $Y=f(x)$ şeklinde alındı. Kimyasal analizlerde kullanılan birçok alet, herhangi bir materyal içinde bulunan bir madde için yapılan analizde, araştırılan maddenin doğrudan oranını değil de orana bağlı olarak ya sayısal bir değer veya bir grafik alanı gösterir (fotometre ve gaz kromatografi cihazı gibi). Bu tip aletler kullanılmadan önce aranılan maddeyi değişik oranlarda içeren çözeltiler hazırlanır. Bunlar sıra ile alete konarak gösterdiği değerler okunur. Burada standart çözeltilerdeki madde oranları bağımsız değişken, aletten okunan değerler ise bağımlı değişkendir. Halbuki araştırmacılar alette okunandan aranan maddenin oranını tahmin etmek isterler. Bu durumda da bundan önceki yaş halkaları örneğinde olduğu gibi, aranan madde oranını

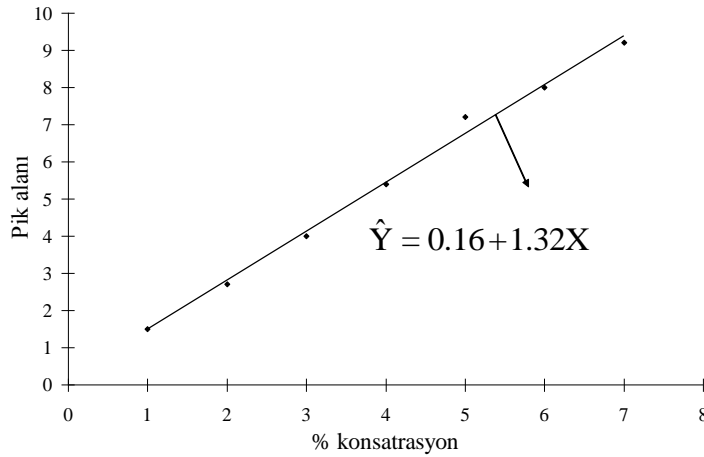
4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

bağımlı değişken almak akla gelir. Ancak kimya alanında aletlerin bu yolla kalibrasyonunda ve aranan madde oranının tahmininde fonksiyon gerçek durumdaki gibi alınagelmektedir. Bağımlı ve bağımsız değişkenin yeri değiştirilmez. Bütün kimya kitaplarında ve laboratuvarlarda böyle yapıldığı için aşağıdaki örnek ele alınmıştır. Üzerinde durulan maddeyi %1, %2, ..., %7 oranında içeren standart çözeltiler (X) hazırlanmış bunlar sıra ile alete konarak pik alanları (Y) okunmuş ve sonuçlar aşağıda özetlenmiştir.

% konsantrasyon X	Pik alanı Y	X ²	Y ²	XY
1	1.5	1	2.25	1.5
2	2.7	4	7.29	5.4
3	4.0	9	16.00	12.0
4	5.4	16	29.16	21.6
5	7.2	25	51.84	36.0
6	8.0	36	64.00	48.0
7	9.2	49	84.64	64.4
Toplam	28	140	255.18	188.9

Noktalar koordinat eksenine işaretlendiğinde (Şekil 5.7) doğrusal bir ilişkinin varlığı görülür.



ŞEKİL 5.7. % konsantrasyonlara karşılık gözlenen pik alanları ve regresyon doğrusu

$\hat{Y} = a + b_{yx}X$ tahmin eşitliğinin katsayıları hesaplanarak denklemde yerine konduğu zaman elde edilen eşitlik aşağıdaki gibidir:

$$\hat{Y} = 0.16 + 1.32X$$

$$r^2 = 0.99$$

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

İçinde aranan maddenin oranının bilinmediği bir çözelti alete konduğunda pik alanı olarak 6.2 okunmuşsa bu değer (Y) yerine konarak X'in aldığı değer çözümlerse aranan oran bulunmuş olur.

$$6.2=0.16+1.32X$$

$$\hat{X} = \frac{6.2 - 0.16}{1.32} \cong 4.58$$

Demek ki üzerinde durulan maddenin oranı ele alınan örnekte %4.58'dir.

Eğer $X=f(y)$ alınsa idi ve tahmin eşitliğinin modeli $\hat{X} = a + b_{yx} Y$ şeklinde olsa idi eşitlik $\hat{X} = -0.1 + 0.755Y$ şeklinde olacaktı. Burada da $r^2=0.99$ 'dur. 6.2 doğrudan yerine konarak;

$$\hat{X} = -0.1 + (0.755)(6.2)$$

$$\hat{X} = 4.58$$

olarak aynı tahmin değeri daha kısa yoldan bulunabilirdi. Yukarıda da değinildiği gibi kimya alanında ilişki olduğu gibi ele alınmaktadır.

5.5. Korelasyon Katsayısına ait Örneklem Dağılımı

X ve Y değişkenleri arasındaki korelasyon katsayısı ρ olan bir populasyondan örnekler çekilse ve bu örneklerde korelasyon katsayıları hesaplanırsa, hesaplanan korelasyon katsayıları örnekten örneğe değişecektir ve bir dağılım gösterecektir. Bu dağılıma "**korelasyon katsayısına ait örneklem dağılımı**" denir.

Korelasyon katsayısına ait örneklem dağılımının parametreleri ortalama, μ_r ve varyans, σ_r^2 'dir ve aşağıdaki şekilde hesaplanır:

$$\mu_r = \rho \text{ ve } \sigma_r^2 = \frac{(1-\rho^2)}{n} \quad \dots(5.7)$$

Varyansın (σ_r^2) hesaplanmasında kullanılan $(1-\rho^2)$ değeri, Y-değerlerinin regresyon denklemi ile açıklanamayan sapmalarından ileri gelen varyasyonun nisbi ölçüsüdür. Eğer $\rho=1.0$ olan bir populasyondan örnekler çekilse ve korelasyon katsayıları hesaplanırsa, hesaplanan korelasyon katsayıları 1'e eşit olacaktır ve bir varyasyon söz konusu olmayacaktır. $(1-\rho^2)$ değeri büyüdükçe (ki bu ρ 'nun küçülmesi demektir) korelasyon katsayılarının gösterdiği varyasyon artar. Aynı zamanda söz konusu varyasyon populasyondan çekilen örneklerin genişliklerine de bağlıdır. Örnek genişliği arttıkça korelasyon katsayıları arasındaki varyasyon azalır.

Eğer popülasyona ait ρ değeri bilinmiyorsa bu durumda korelasyon katsayısına ait örneklem dağılımının varyansı hesaplanırken örnekten hesaplanan değer kullanılır ve varyans aşağıdaki şekilde hesaplanır:

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

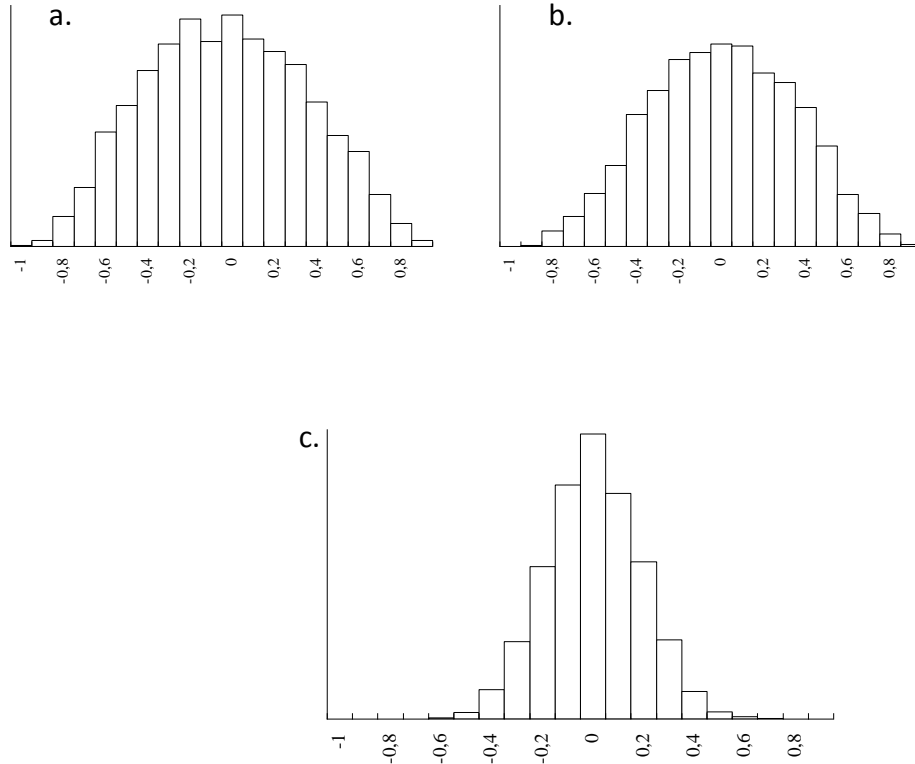
$$S_r^2 = \frac{(1-r^2)}{(n-2)} \quad \dots(5.8)$$

Korelasyon katsayısının örnekten hesaplanan standart sapması ise aşağıdaki gibidir:

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

Daha önce de başka istatistikler için belirtildiği gibi buna kısaca korelasyon katsayısının **standart hatası** denir.

Simulasyon yöntemi ile $\rho=0$ olan bir populasyondan örnek genişliği $n=8$, $n=10$ ve $n=30$ yani bu sayılarda X ve Y değişken çifti içeren çok sayıda örnekler çekilmiş ve bunlardan korelasyon katsayısı hesaplanmıştır. Her örnek genişliğinde de korelasyon katsayılarının çoğunluğu sıfır etrafında toplanmıştır. Bir kısmı pozitif bir kısmı negatiftir. Ortalamaları sıfırdır. Bunların histogramları yapıldığında Şekil 5.8'de görüldüğü gibi örneklerden hesaplanan korelasyon katsayıları normal dağılıma yaklaşmaktadır. Bu normal dağılımların standart sapmaları örnek genişliği arttıkça azalmaktadır.

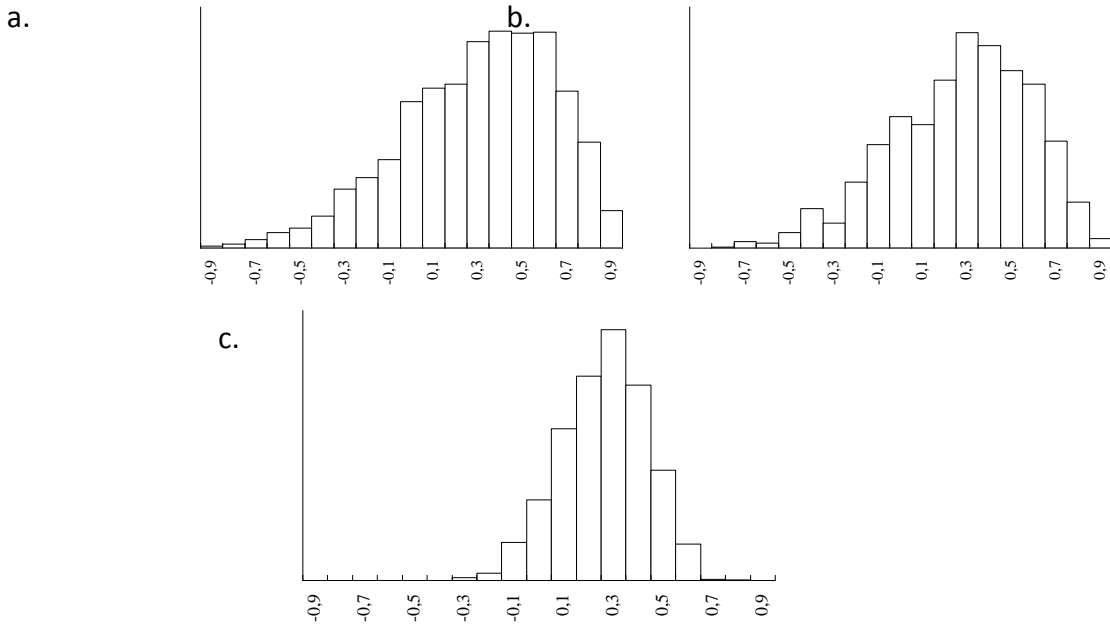


ŞEKİL 5.8. Simulasyon yöntemi ile üretilmiş ve $\rho=0$ olan populasyondan çekilmiş ve a.) $n=8$, b.) $n=10$ ve c.) $n=30$ birey içeren örneklerden elde edilmiş korelasyon katsayısına ait histogramlar

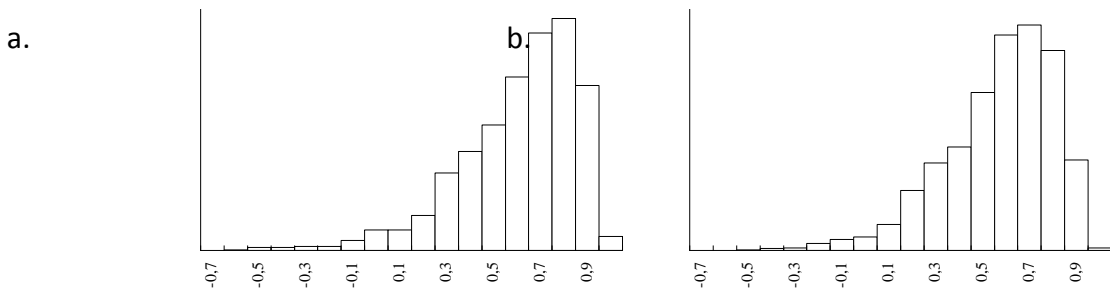
4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Populasyona ait korelasyon katsayısı (ρ) sıfırdan farklılaştıkça, örnekten hesaplanan korelasyon katsayısının (r) dağılım şeklinin simetriden uzaklaştığı Şekil 5.9, 5.10 ve 5.11’de görülmektedir. Şekil 5.9.a., b. ve c., simülasyon yöntemi ile üretilmiş ve $\rho=0.3$ olan populasyondan çekilmiş 8, 10 ve 30 birey içeren örneklerden elde edilmiş korelasyon katsayısına ait örnekleme dağılımlarını göstermektedir. Şekil 5.10.a., b. ve c., simülasyon yöntemi ile üretilmiş ve $\rho=0.6$ olan populasyondan çekilmiş 8, 10 ve 30 birey içeren örneklerden elde edilmiş korelasyon katsayısına ait örnekleme dağılımlarını göstermektedir. Şekil 5.11.a., b. ve c. ise simülasyon yöntemi ile üretilmiş ve $\rho=0.9$ olan populasyondan çekilmiş 8, 10 ve 30 birey içeren örneklerden elde edilmiş korelasyon katsayısına ait örnekleme dağılımlarını göstermektedir. Şekil 5.9, 5.10 ve 5.11’de verilen korelasyona ait örnekleme dağılımlarından görülebileceği gibi, bu populasyonlardan çekilen örneklerin genişliği arttıkça dağılım normale biraz yaklaşır.

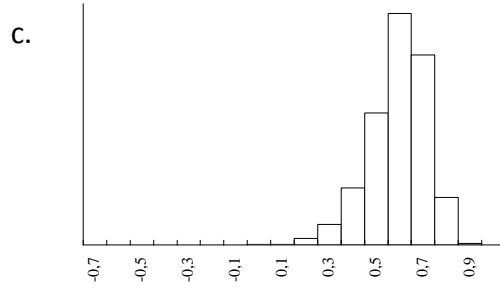


ŞEKİL 5.9. Simülasyon yöntemi ile üretilmiş ve $\rho=0.3$ olan populasyondan çekilmiş ve a.) $n=8$, b.) $n=10$ ve c.) $n=30$ birey içeren örneklerden elde edilmiş korelasyon katsayısına ait histogramlar

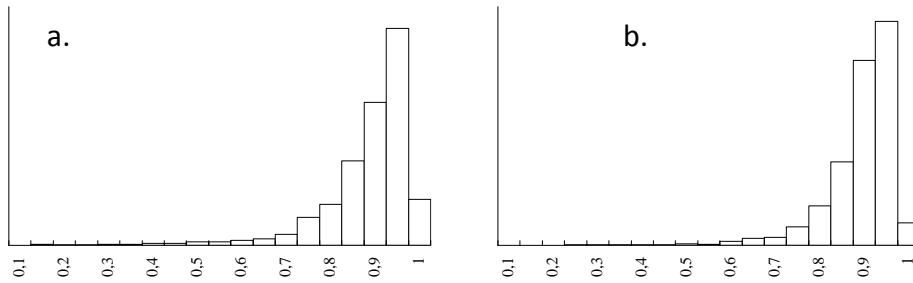


4. HAFTA DERS NOTLARI

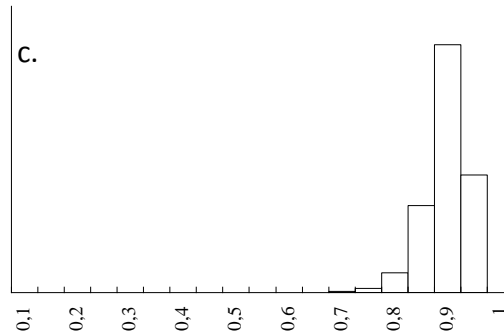
Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94



ŞEKİL 5.10. Simulasyon yöntemi ile üretilmiş ve $\rho=0.6$ olan populasyondan çekilmiş ve a.) $n=8$, b.) $n=10$ ve c.) $n=30$ birey çifti içeren örneklerden elde edilmiş korelasyon katsayısına ait histogramlar



ŞEKİL 5.11. Simulasyon yöntemi ile üretilmiş ve $\rho=0.9$ olan populasyondan çekilmiş ve a.) $n=8$ ve b.) $n=10$ birey içeren örneklerden elde edilmiş korelasyon katsayısına ait histogramlar



ŞEKİL 5.11.(devam). Simulasyon yöntemi ile üretilmiş ve $\rho=0.9$ olan populasyondan çekilmiş c.) $n=30$ birey içeren örneklerden elde edilmiş korelasyon katsayısına ait histogramlar

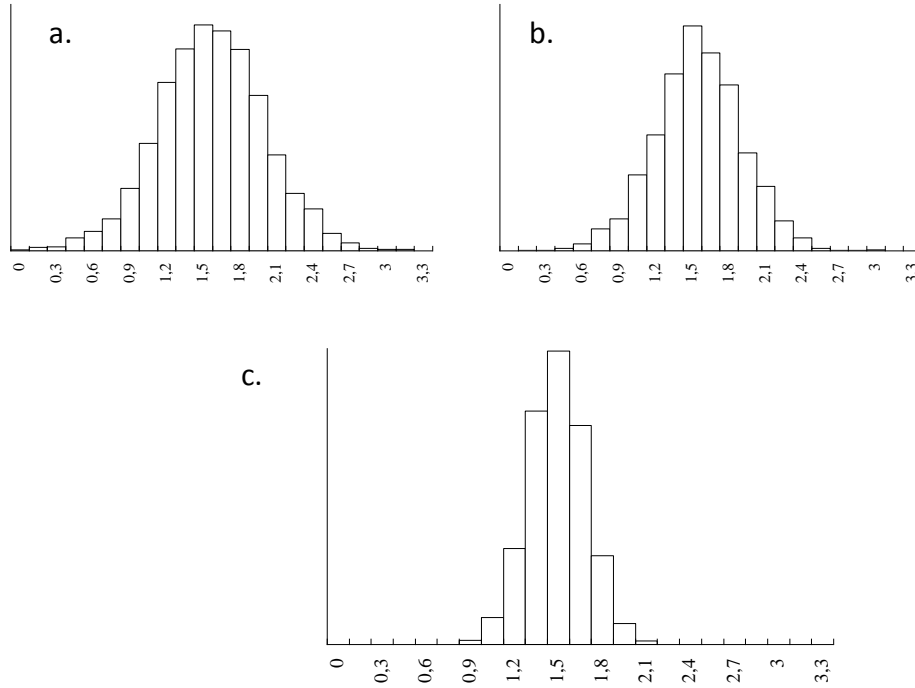
Populasyonda korelasyon katsayısının (ρ) sıfırdan farklı olduğu durumlarda örneklerden hesaplanan r değerleri aşağıdaki şekilde Z_r -değerlerine dönüştürülürse bu değerler yaklaşık normal dağılır.

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

$$Z_r = \frac{1}{2} \log_e \frac{(1+r)}{(1-r)} = 1.1513 \log \frac{(1+r)}{(1-r)} \quad \dots(5.9)$$

Elde edilen Z_r -değerleri ρ ne olursa olsun normal dağılıma yaklaşır. Şekil 5.12’de örnek olarak $\rho=0.9$ olan populasyondan çekilmiş 8, 10 ve 30 birey içeren örneklerden elde edilen ve dağılım şekli Şekil 5.11’de verilen örnekleme dağılımını oluşturan r değerleri için hesaplanmış Z_r -değerlerinin dağılımı verilmektedir. Şekil 5.12’de görüldüğü gibi $\rho=0.9$ olmasına rağmen örnek genişliği ne olursa olsun Z_r -değerlerinin dağılımı normale yaklaşmaktadır. Aynı şekilde $\rho=0.3$ ve $\rho=0.6$ olan populasyonlardan çekilmiş örneklerden elde edilen korelasyon katsayılarına karşılık gelen Z_r -değerleri de hesaplanırsa, bu hesaplanan değerler de normal dağılıma yaklaşır.



ŞEKİL 5.12. Simulasyon yöntemi ile üretilmiş ve $\rho=0.9$ olan populasyondan çekilmiş ve a.) $n=8$, b.) $n=10$ ve c.) $n=30$ birey içeren örneklerden elde edilmiş korelasyon katsayıları için hesaplanmış Z_r -değerlerine ait histogramlar

Bu Z_r -değerlerinin ortalaması;

$$\mu_{Z_r} = \frac{1}{2} \log_e \frac{(1+\rho)}{(1-\rho)} \quad \dots(5.10)$$

ve standart sapması;

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

$$\sigma_{Z_r} = \sqrt{\frac{1}{(n-3)}} \quad \dots(5.11)$$

tür.

Korelasyon katsayısı ρ olan bir populasyondan n_1 ve n_2 örnekleme genişliğinde örnekler çekilse ve bu örneklerde korelasyon katsayıları hesaplanırsa (r_1 ve r_2 'ler) ve bunlar tamamen tesadüfen yan yana getirilerek farkları alınsa bu farklar bir dağılım gösterir ki bu dağılıma “**korelasyon katsayıları arasındaki farka ait örnekleme dağılımı**” denir. Korelasyon katsayılarına karşılık hesaplanacak Z_r -değerleri farkı bir normal dağılıma yaklaşır. Bu farkların gösterdiği dağılımın ortalaması;

$$\mu_{(Z_{r1}-Z_{r2})} = 0$$

ve standart sapması;

$$\sigma_{(Z_{r1}-Z_{r2})} = \sqrt{\frac{1}{(n_1-3)} + \frac{1}{(n_2-3)}} \quad \dots(5.12)$$

tür.

5.6. Regresyon Katsayısına ait Örnekleme Dağılımı

Regresyon katsayısı β_{yx} olan populasyondan n birey içeren örnekler çekilse ve regresyon katsayıları hesaplanırsa, hesaplanan regresyon katsayıları (b_{yx} 'ler) büyük n değerleri için normal dağılıma yaklaşır. Bu dağılıma regresyon katsayısına ait örnekleme dağılımı denir. Bu dağılımın ortalaması, β_{yx} ve varyansı;

$$\sigma_b^2 = \frac{\sigma_e^2}{\sum d_x^2} \quad \dots(5.13)$$

dir.

5.13 numaralı eşitlikteki σ_e^2 , $Y = \alpha + \beta X + e$ modelinin α ve β parametreleri bilindiğinde $\sigma_e^2 = \frac{\sum e^2}{n} = \frac{\sum (Y - \hat{Y})^2}{n}$ şeklinde hesaplanır. Burada $\sum e^2$, Şekil 5.6'da görüldüğü gibi, gözlemlerin regresyon doğrusundan sapmalarının karelerinin toplamı olup aşağıdaki şekilde hesaplanır;

$$\sum (Y - \hat{Y})^2 = \sum e^2 = \sum d_y^2 - \beta_{yx} \sum d_x d_y$$

β_{yx} 'in eşiti olan $\frac{\sum d_x d_y}{\sum d_x^2}$ yerine konduğunda daha kısa olarak aşağıdaki eşitlikten

de hesaplanır:

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

$$\sum e^2 = \sum d_y^2 - \frac{(\sum d_x d_y)^2}{\sum d_x^2}$$

Eğer populasyona ait regresyon eşitliğinin katsayıları bilinmiyor ve örnekten tahmin ediliyorsa bu durumda regresyon katsayısına ait örnekleme dağılımının varyansı;

$$S_b^2 = \frac{S_e^2}{\sum d_x^2} \quad \dots(5.14)$$

dir. Regresyon katsayısının standart hatası ise aşağıdaki gibidir:

$$S_b = \sqrt{\frac{S_e^2}{\sum d_x^2}}$$

Eşitlikte, S_e^2 regresyon doğrusundan sapma kareler ortalamasıdır ve aşağıdaki şekilde hesaplanır:

$$S_e^2 = \frac{\sum (Y - \hat{Y})^2}{n - 2} = \frac{\sum d_y^2 - b_{yx} \sum d_x d_y}{n - 2} = \frac{\sum d_y^2 - \frac{(\sum d_x d_y)^2}{\sum d_x^2}}{n - 2} \quad \dots(5.15)$$

Regresyondan sapma kareler toplamı (n-2)'ye bölünerek S_e^2 bulunmuştur. Bunun nedeni regresyon eşitliğinde α ve β parametreleri yerine bunların örnekten tahminlerinin yani istatistiklerin kullanılmış olmasıdır.

Regresyon eşitliğinin diğer katsayısı olan a'nın standart hatasının eşiti ise aşağıdaki gibidir:

$$S_a = \sqrt{S_e^2 \frac{\sum X_i^2}{n \sum d_x^2}} \quad \text{veya} \quad S_a = S_e \sqrt{\frac{\sum X_i^2}{n \sum d_x^2}} \quad \dots(5.16)$$

$\hat{Y}_i = a + b_{yx} X_i$ eşitliğinden herhangi bir X_i noktası için \hat{Y}_i tahmini yapılmış ise bunun standart hatası ($S_{\hat{y}_i}$) aşağıdaki gibidir:

$$S_{\hat{y}_i} = S_e \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum d_x^2}} \quad \dots(5.17)$$

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Eğer örnek genişliği yeteri kadar fazla ise kök içindeki $\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum d_x^2}$ ihmal edilecek kadar küçük olur. Bu durumda tahminin standart hatası olarak S_e alınır. Yani $S_{\hat{y}_i} = S_e$ olarak alınabilir.

ÖRNEK:

Örnek 1'de haftalar boyunca (X) C-vitamininin (Y) değişimi incelenmişti. Regresyon eşitliği $\hat{Y} = 104 - 9.5X$ olarak bulunmuştu. Bu örnek için aşağıdaki sonuçlar hesaplanabilir.

$$\sum (Y_i - \hat{Y})^2 = \sum e^2 = 3720 - \frac{(-380)^2}{40} = 110.0$$

$$S_e^2 = \frac{110.0}{(5-2)} = 36.67 \quad S_e = \sqrt{36.67} = 6.055$$

$$S_b = \sqrt{\frac{36.67}{40}} = 0.957 \quad S_a = \sqrt{36.67 \frac{120}{(5)(40)}} = 4.69$$

Yaş halkaları genişliği ile yıllık yağış miktarı arasındaki örnek için regresyon eşitliği $\hat{Y} = 203.14 + 77.15X$ olarak bulunmuştu. Ele alınan örnekteki sonuçlara göre:

$$\sum (Y_i - \hat{Y}_i)^2 = \sum e^2 = 7716.1 - \frac{(80.31)^2}{1.041}$$

$$\sum e^2 = 1520.4$$

$$S_e^2 = \frac{1520.4}{10-2} = 190.5 \quad S_e = \sqrt{190.5} = 13.79$$

X=3.8 mm yaş halkası genişliği için $\hat{Y} = 496.31$ mm yağış tahmin edilmişti. Bu tahminin standart hatası;

$$S_{\hat{y}} = 13.79 \sqrt{1 + \frac{1}{10} + \frac{(3.8 - 3.17)^2}{1.041}} = 16.78$$

Buna göre 3.8 mm halka genişliği için yapılan yağış miktarı tahmini 496.31 ± 16.78 mm olarak belirtilir. Daha önce de belirtildiği gibi tahminler standart hataları ile birlikte verilirse bir bilimsel değer taşır.

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Regresyon için verilen Örnek 3’de pik alanı için eşitlik $\hat{Y} = 0.16 + 1.32X$ olarak belirlenmişti. Kimya laboratuvarlarında aletten okunan Y değerleri yerine konarak X değerinin (konsantrasyonun) tahmin edildiği açıklanmıştı. Böyle tahmin edilen bir \hat{x} değerinin (konsantrasyonun) standart hatası ise aşağıdaki gibidir:

$$S_{\hat{x}_i} = \frac{S_e}{b_{yx}} \sqrt{1 + \frac{1}{n} + \frac{(y_i - \bar{y})^2}{b_{yx}^2 \sum d_x^2}} \quad \dots(5.18)$$

Örnek 3 için;

$$\bar{X} = 4.0 \quad \bar{Y} = 5.43$$

$$\sum d_x d_y = 36.9 \quad \sum d_x^2 = 28.0 \quad \sum d_y^2 = 48.89$$

değerleri hesaplanırsa, $Y=6.2$ pik alanı için hesaplanan $\hat{x} = 4.58$ konsantrasyonunun standart hatası aşağıdaki gibi hesaplanır:

$$S_{\hat{x}} = \frac{0.229}{1.32} \sqrt{1 + \frac{1}{7} + \frac{(6.2 - 5.43)^2}{(1.32)^2 (28)}} = 0.186$$

Tahmin edilen konsantrasyon standart hatası ile birlikte 4.58 ± 0.186 olarak gösterilir.

Regresyon katsayısı β_{yx} olan populasyondan n birey içeren örnekler çekilse ve regresyon katsayıları hesaplanırsa, hesaplanan regresyon katsayıları (b_{yx} ’ler) tamamen tesadüfen yan yana getirilerek farkları alınsa bu farklar bir dağılım gösterir ki bu dağılıma regresyon katsayıları arasındaki farka ait örnekleme dağılımı denir. Bu dağılımın ortalaması, $\mu_{(b_1-b_2)}=0$ ve varyansı;

$$\sigma_{(b_1-b_2)}^2 = \sigma_{b_1}^2 + \sigma_{b_2}^2$$

$$S_{(b_1-b_2)}^2 = \frac{S_e^2}{\sum d_{x1}^2} + \frac{S_e^2}{\sum d_{x2}^2}$$

Örnekler aynı populasyondan çekildiği için σ_e^2 ’ler aynıdır ve eşitlik aşağıdaki şekilde yazılabilir:

$$\sigma_{(b_1-b_2)}^2 = \sigma_e^2 \left(\frac{1}{\sum d_{x1}^2} + \frac{1}{\sum d_{x2}^2} \right) \quad \dots(5.19)$$

4. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Regresyon katsayıları arasındaki farka ait örnekleme dağılımının standart sapması ise:

$$\sigma_{(b_1-b_2)} = \sqrt{\sigma_e^2 \left(\frac{1}{\sum d_{x_1}^2} + \frac{1}{\sum d_{x_2}^2} \right)} \quad \dots(5.20)$$

Populasyona ait regresyon katsayısı bilinmiyor ve σ_e^2 örnekten tahmin ediliyorsa S_e^2 örneklerden hesaplanan regresyondan sapma kareler ortalamalarının serbestlik dereceleri ile tartılı ortalamasıdır ve aşağıdaki şekilde hesaplanır:

$$S_e^2 = \frac{\sum (Y_1 - \hat{Y}_1)^2 + \sum (Y_2 - \hat{Y}_2)^2}{(n_1 - 2) + (n_2 - 2)} \quad \dots(5.21)$$

ve regresyon katsayıları arasındaki farka ait dağılımın standart sapması ise 5.22 numaralı eşitlikle verildiği gibidir.

$$S_{(b_1-b_2)} = \sqrt{S_e^2 \left(\frac{1}{\sum d_{x_1}^2} + \frac{1}{\sum d_{x_2}^2} \right)} \quad \dots(5.22)$$

dir.