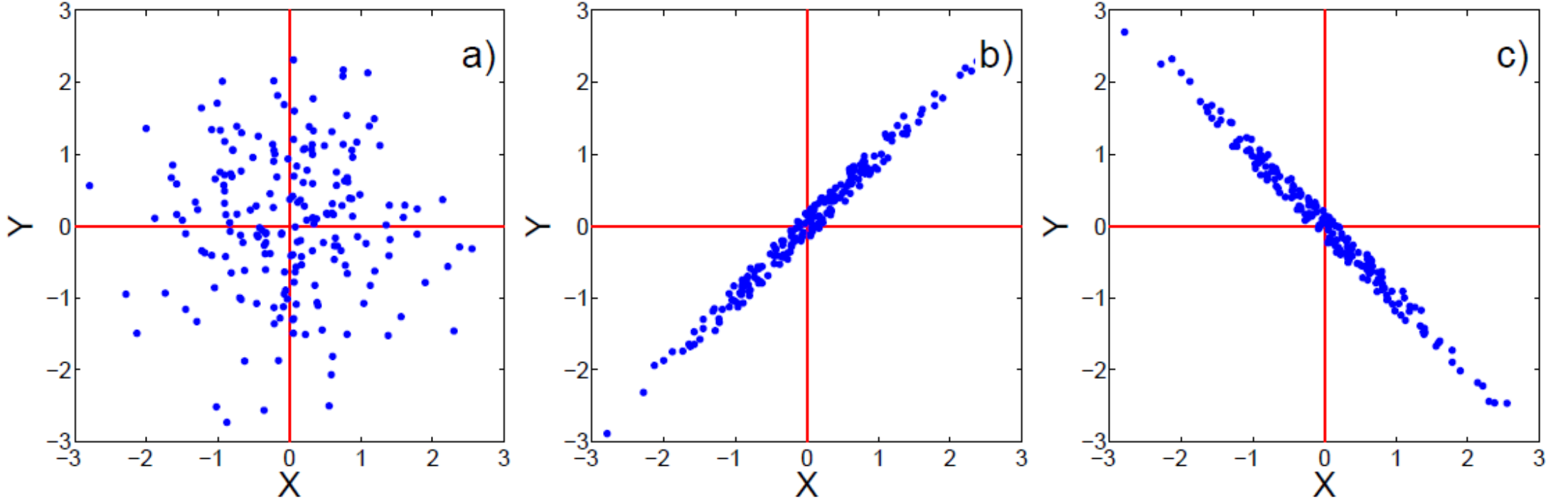


İSTATİKSEL İLİŞKİLER

REGRESYON VE KORELASYON

İlişkinin Yönü Derecesi, Fonksiyonel Şekli ve Korelasyon Tipleri



Regresyon: ilişkinin fonksiyonel tipini ve mutlak ölçüler içinde bağılılığın şekli

Korelasyon: ilişkinin derecesi hakkında oransal bir ölçü

İlişkinin Derecesini Belirlemede Kullanılan Ölçüler

- Bağılılık indeksi
- Bağılılık katsayısı
- Birlikte değişim katsayısı
- Korelasyon katsayısı

- ilk üçü değerlerin birbirlerine göre, sonuncusu yani korelasyon katsayısı ise ortalamaya göre sapmalarına dayanır.

Bağılılık İndeksi

- Değişmelerin yalnız işaretlerine dayanan basit bir ilişki ölçüsüdür.

$$B \cdot i = \frac{D_p - D_n}{N - 1}$$

$B \cdot i$ = Bağılılık indeksi

D_p = İki seride işaretleri birbirine uyan değişmelerin sayısı

D_n = İki seride işaretleri birbirine uymayan değişmelerin sayısı

$N-1$ = Bütün değişmelerin sayısı (serilerde değer sayısı N olduğuna göre değişmeler sayısı bundan 1 eksik yani $N-1$ 'dir).

$$D_p + D_n = N - 1$$

- Bağılılık indeksi (-1) ile (+1) arasında değişen bir sonuç verir.
- (+1) çıkması ilişkinin tam ve aynı yönde , (-1) bulunması ise ilişkinin tam ve ters olduğunu gösterir.
- Sonucun 0 çıkması iki değişken arasında hiçbir ilişki bulunmadığını gösterir.

Örnek: Türkiye’de fındık ağacı sayısı ile üretim arasındaki ilişkinin derecesini bağıllık indeksi ile tespit edelim. Üretim ağaç sayısına bağlı olarak değişeceğine göre ağaç sayısı serbest değişken (x), üretim ise bağlı değişken (y) sayılması gerekir.

Yıllar	Ağaç Sayısı (bin)	Üretim (bin ton)	Değişmeler		Karşılıklı Değişmelerde	
	x	y	Δx	Δy	Uyan	Uymayan
1960	161	58				
1961	169	76	+	+	+	
1962	175	122	+	+	+	
1963	176	88	+	-		-
1964	184	195	+	+	+	
1965	187	62	+	-		-
1966	192	190	+	+	+	
1967	195	71	+	-		-
1968	197	138	+	+	+	
1969	194	170	-	+		-
					$D_p = 5$	$D_n = 4$

$$B \cdot i = \frac{D_p - D_n}{N - 1} = \frac{D_p - D_n}{D_p + D_n} = \frac{5 - 4}{5 + 4} = 0.11$$

Görüldüğü gibi bağıllık indeksine göre fındık üretimi ile ağaç sayısı arasında doğrusal (müspet), fakat çok zayıf bir ilişki bulunduğu sonucuna varılmaktadır.

Bağılılık İndeksinin Dezavantajı

Örnek: x ve y değerleri aşağıdaki şekilde olmak üzere bağılılık indeksi

x	y	Δx	Δy	Uyan	Uymayan
30	30				
20	20	-	-	+	
21	19	+	-		-
				$D_p = 1$	$D_n = 1$

$$B \cdot i = \frac{D_p - D_n}{N - 1} = \frac{1 - 1}{3 + 1} = 0$$

Bağılılık Katsayısı

- Değişmelerin yalnız işaretlerini değil büyüklüklerini de hesaba katar.

$$B \cdot K = \frac{\sum (\Delta x \cdot \Delta y)}{\sum |\Delta x \cdot \Delta y|}$$

Örnek: Fındık konusunda ağaç sayısı ile üretim miktarı arasındaki ilişkiyi şimdi bağılılık katsayısı ile hesaplayalım.

Ağaç Üretim
Sayısı (bin) (bin ton)

Yıllar	x	y	Δx	Δy	$\Delta x \cdot \Delta y$
1960	161	58			
1961	169	76	+8	+18	144
1962	175	122	+6	+46	276
1963	176	88	+1	-34	-34
1964	184	195	+8	+107	856
1965	187	62	+3	-133	-399
1966	192	190	+5	+128	640
1967	195	71	+3	-119	-357
1968	197	138	+2	+67	134
1969	194	170	-3	+32	-96
					2050 -886

$$B \cdot K = \frac{\sum (\Delta x \cdot \Delta y)}{\sum |\Delta x \cdot \Delta y|} = \frac{2050 - 886}{2050 + 886} = \frac{1164}{2936} = 0.40$$

- **Bağılılık Katsayısının Dezavantajları**

Örnek:

x	y	Δx	Δy	$\Delta x \cdot \Delta y$
10	5			
20	10	+10	+5	+50
20	15	0	+5	0
30	20	+10	+5	+50
				100

$$B \cdot K = \frac{\sum (\Delta x \cdot \Delta y)}{\sum |\Delta x \cdot \Delta y|} = \frac{100}{100} = 1$$

Birlikte Değişim Katsayısı

Tam ve pozitif ilişki halinde

x	y	Δx	Δy
5	12		
8	15	+3	+3
10	17	+2	+2
6	13	-4	-4
5	12	-1	-1
10	17	+5	+5

Tam ve ters ilişki halinde

x	y	Δx	Δy
5	12		
8	9	+3	-3
10	7	+2	-2
6	11	-4	+4
5	12	-1	+1
10	7	+5	-5

$$\sum(\Delta x)^2 = \sum(\Delta y)^2 = \sum|\Delta x \cdot \Delta y|$$

$$B \cdot D \cdot K = \frac{\sum(\Delta x \cdot \Delta y)}{\sqrt{\sum(\Delta x)^2 \cdot \sum(\Delta y)^2}}$$

- Tam ilişki halinde oranın payı paydasına eşit olacağından sonuç 1 çıkar.
- (+1) tam ve pozitif yönde bir ilişkiyi, (-1) ise tam ve negatif yönde bir ilişkiyi gösterir.
- Sonucun sıfır çıkması iki değişken arasında her hangi bir ilişkinin bulunmadığını gösterir.

Örnek: 1960-1969 yılları arasında fındık ağaçları sayısı ile üretim arasındaki ilişki bu defa birlikte değişim katsayısı ile hesaplayalım.

	Ağaç Sayısı (bin)	Üretim (bin ton)					
Yıllar	x	y	Δx	Δy	$\Delta x \cdot \Delta y$	$(\Delta x)^2$	$(\Delta y)^2$
1960	161	58	-	-	-	-	-
1961	169	76	+8	+18	144	64	324
1962	175	122	+6	+46	276	36	2116
1963	176	88	+1	-34	-34	1	1156
1964	184	195	+8	+107	856	64	11449
1965	187	62	+3	-133	-399	9	17689
1966	192	190	+5	+128	640	25	16384
1967	195	71	+3	-119	-357	9	14161
1968	197	138	+2	+67	134	4	4489
1969	194	170	-3	+32	-96	9	1024
					2050 -886	221	68792

$$B \cdot D \cdot K = \frac{\sum (\Delta x \cdot \Delta y)}{\sqrt{\sum (\Delta x)^2 \cdot \sum (\Delta y)^2}} = \frac{2050 - 886}{\sqrt{221 \cdot 68792}} = 0.30$$

Korelasyon Katsayısı

- iki seride değerlerin ortalamadan farklarını (sapmalarını) esas alır ve değişimleri bunlara göre inceler

Δx 'ler yerine $X_i - \bar{x} = x_i$

Δy 'ler yerine $Y_i - \bar{y} = y_i$

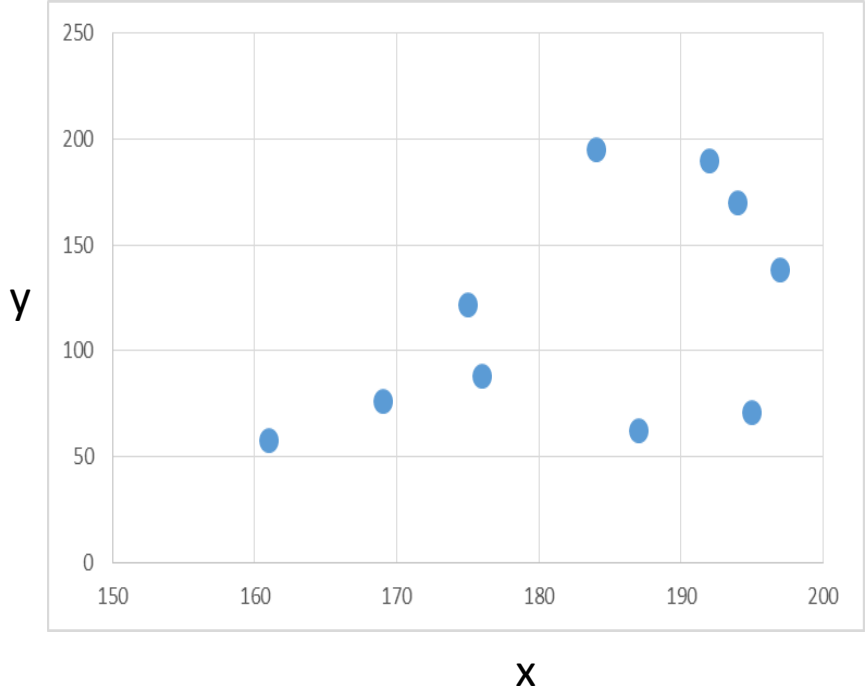
- Korelasyon katsayısı

$$r = \frac{\sum (X_i - \bar{x}) \cdot (Y_i - \bar{y})}{\sqrt{\sum (X_i - \bar{x})^2 \sum (Y_i - \bar{y})^2}}$$

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

Örnek: Fındık konusunda ağaç sayısı ile üretim arasındaki ilişkiyi son olarak korelasyon katsayısıyla hesaplayalım. Gerekli hesap işlemleri aşağıdaki tabloda verilmiştir.

	Ağaç Sayısı (bin)	Üretim (bin ton)					
Yıllar	x	y	$x_i = X_i - \bar{x}$	$y_i = Y_i - \bar{y}$	$x_i \cdot y_i$	x_i^2	y_i^2
1960	161	58	-22	-59	1298	484	3481
1961	169	76	-14	-41	574	196	1681
1962	175	122	-8	+5	-40	64	25
1963	176	88	-7	-29	203	49	841
1964	184	195	+1	+78	78	1	6084
1965	187	62	+4	-55	-220	16	3025
1966	192	190	+9	+73	657	81	5329
1967	195	71	+12	-46	-552	144	2116
1968	197	138	+14	+21	294	196	441
1969	194	170	+11	+53	583	121	2809
	$\sum x = 1830$	$\sum y = 1170$			2875	1352	25832
	$\bar{x} = 183$	$\bar{y} = 117$					



$$r = \frac{\sum x_i \cdot y_i}{\sqrt{\sum x_i^2 \cdot \sum y_i^2}} = \frac{2875}{\sqrt{1352 \cdot 25832}} = 0.49$$

Aynı örnek için
bağıllık indeksi: 0.11,
bağıllık katsayısı: 0.40,
birlikte değişim katsayısı: 0.30
korelasyon katsayısı: 0.49 çıkmıştır.

Korelasyon katsayısı yalnız ortalamadan sapmalara göre değil, esas değerlere göre de hesaplanabilir ve şu formül kullanılır.

$$r = \frac{n \cdot \sum_{i=1}^n x_i Y_i - \sum_{i=1}^n x_i \sum_{i=1}^n Y_i}{\sqrt{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \left(n \sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i \right)^2 \right)}}$$

Ayrıca Korelasyon katsayısının basit formülünde paydaya yazılan değeri veri sayısı olan n ile bir defa çarpar, bir defa bölersek değeri değişmeyeceğinden formül şu şekilde yazılabilir.

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

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

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

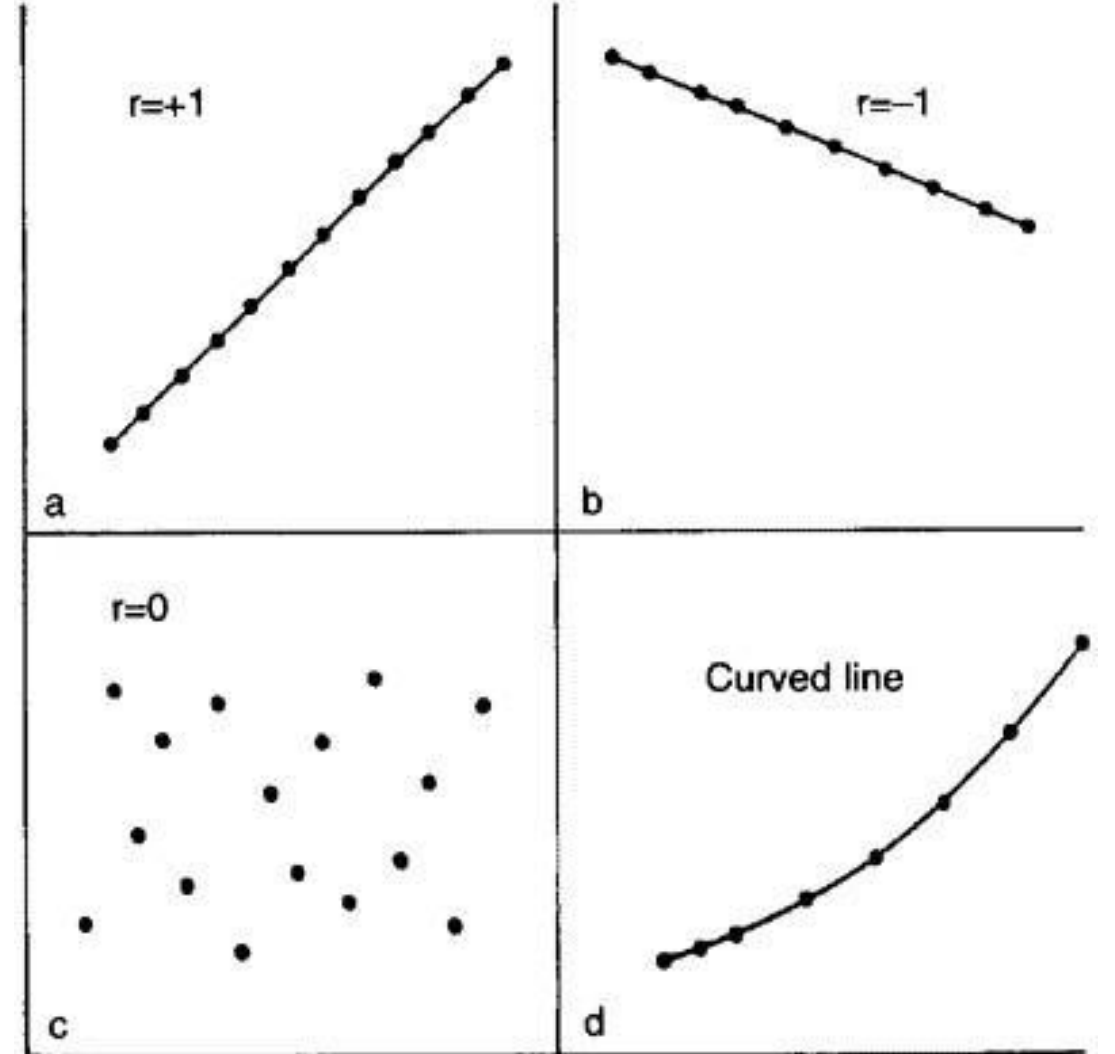
$$r = \frac{\sum x_i \cdot y_i}{n \cdot \sigma_x \cdot \sigma_y}$$

Korelasyon Katsayısının Özellikleri

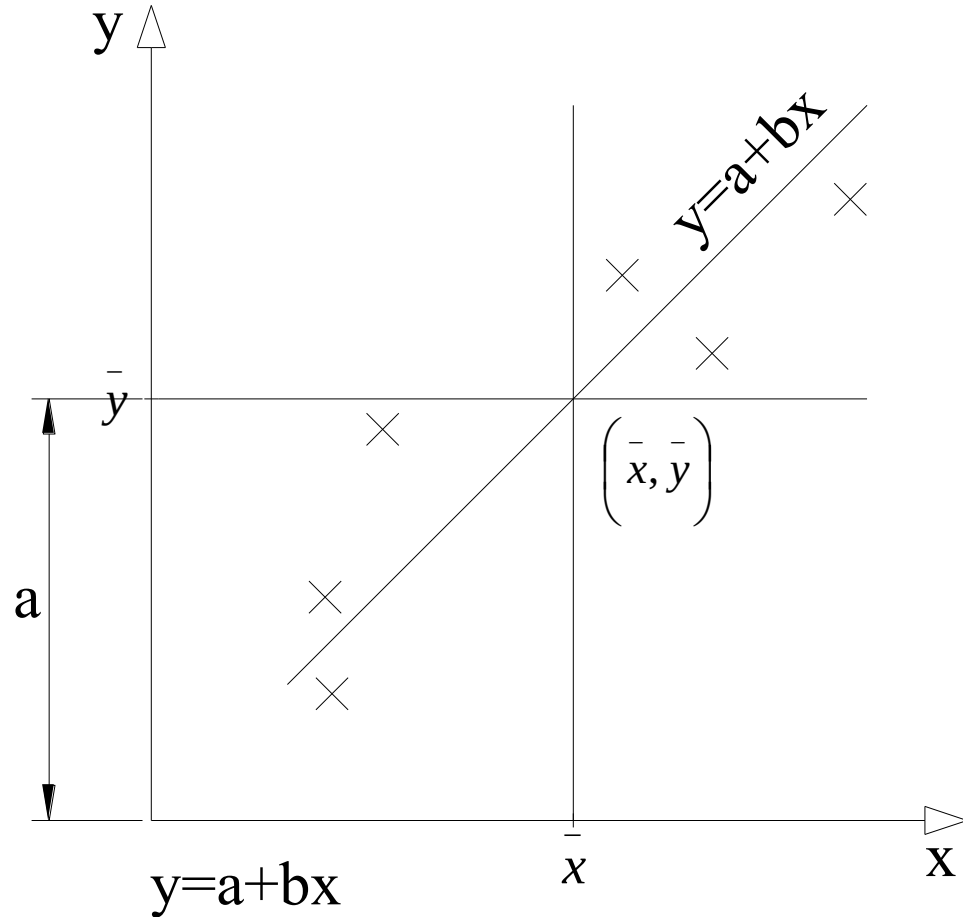
1. Bilimsel araştırmalarda en sağlıklı sonuçları verdiği için korelasyon katsayısı uygulamada en yaygın şekilde kullanılan ilişki ölçüsüdür.
2. Seride anormal (aşırı) değerlerin varlığı korelasyon katsayısına büyük ölçüde etki eder. Böyle hallerde korelasyon katsayısı gerçeği göstermeyeceğinden mümkünse bu anormal değerlerin giderilmesi (atılması) gerekir.
3. Korelasyon katsayısı değerlendirilirken hesaba giren veri sayısı (n) önemlidir. Tesadüfi sebeplerin tesiri daha iyi giderileceği için veri sayısı çoğaldıkça bulunacak sonuca daha çok güvenilebilir.
4. Seri arasında mantıklı bir ilişki olup olmadığı noktası üzerinde önemle durulması gerekir. Bazen aralarında mantıksal bir ilişki bulunmayan seriler için de tesadüfen yüksek bir korelasyon katsayısı elde edilebilir.

Basit (Doğrusal) Regresyon ve Korelasyon

- Regresyon: iki değişken (seri) arasındaki ortalama bir ilişkinin matematiksel ifadesi
- Korelasyon: ilişkinin derecesinin tayini



Regresyon



$$b = \frac{\sum_{i=1}^n (X_i - \bar{x})(Y_i - \bar{y})}{\sum_{i=1}^n (X_i - \bar{x})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$b = \frac{\sigma_y}{\sigma_x} \cdot r$$

$$a = \bar{y} - b\bar{x}$$

Örnek: Dicle nehrinin Rezuk akım ölçme istasyonunun Sinan istasyonunda ölçülen yıllık akış hacmi aşağıda verilmiştir.

Yıl	R (10^6 m^3)	S (10^6 m^3)
1956	16077	4629
1957	14867	4556
1958	11720	2507
1959	9352	1612
1960	10537	2125
1961	6743	1054
1962	10162	2272
1963	29232	11883
1964	16439	?
1965	13019	4041
1966	17729	5191
1967	20368	5328
1968	26748	6543
1969	33566	7606
1970	12314	3445
1971	10914	3161

- İki istasyondaki akışlar arasındaki korelasyon katsayısını hesaplayınız.
- Sinan'daki akışların Rezuk' taki akışlara göre regresyon doğrusunun denklemini elde ediniz.
- 1664 yılında ölçülmemiş olan Sinan istasyonunun yıllık akış hacmini tahmin ediniz.

Çözüm:

$$\bar{x} = 1637 \cdot 10^6 \text{ m}^3 \quad \sigma_x = 7669 \cdot 10^6 \text{ m}^3 \quad \sum x_i y_i = 137 \cdot 10^{18} \text{ m}^3$$
$$\bar{y} = 4397 \cdot 10^6 \text{ m}^3 \quad \sigma_y = 2768 \cdot 10^6 \text{ m}^3$$

a. Bu değerleri kullanarak korelasyon katsayısı, $r = 0.868$ bulunur.

b. $\bar{y} = a + b\bar{x}$

$$b = \frac{\sigma_y}{\sigma_x} \cdot r = \frac{2768 \cdot 10^6}{7669 \cdot 10^6} \cdot 0.868 = 0.313$$

$$a = \bar{y} - b\bar{x} = 4397 \cdot 10^6 - 0.313 \cdot 16233 \cdot 10^6$$

$$a = -684 \cdot 10^6$$

$$\bar{y} = -684 \cdot 10^6 + 0.313x$$

c. 1964 yılında ölçülmemiş olan Sinan istasyonundaki akışın tahmini,

$$\bar{y} = -684 \cdot 10^6 + 0.313 \cdot 16439 \cdot 10^6 = 4461 \cdot 10^6 \text{ m}^3$$

olarak bulunur.

→ dışında $|z| \geq 1.96$ veya $|z| \geq 2.58$ olup olmadığı araştırılır.

