

Daha ileriye... En iyiye...

50.
Yıl



**HACETTEPE
ÜNİVERSİTESİ**

www.hacettepe.edu.tr

KORELASYON

- İki değişken arasındaki ilişkiyi tanımlamak ve ölçmek için kullanılan istatistiksel yöntem korelasyon denir.
- Genellikle iki değişken doğal ortamında ölçülür.
- İlgilendiğimiz şey bir değişkenin (X) diğer bir değişkenle (Y) ilişkili olup olmadığıdır.
- Herhangi bir manipulasyon yapmayız, kontrol etmeye çalışmayız.
- Değişkenleri bağımlı veya bağımsız olarak sınıflandırmayız.

Örnek araştırma soruları

Deneysel desenler	Korelasyonel desenler
<ul style="list-style-type: none">• A öğretim yönteminin öğrencilerin matematik performansları üzerindeki etkisi nedir? <p>Bağımlı Değişken= Matematik performansı</p> <p>Bağımsız değişken= Öğretim yöntemi</p> <ul style="list-style-type: none">• Örneklem verisi deneyin/tedavinin anlamlı bir etkisi olduğunu gösterecek bir kanıt sunmakta mıdır?	<ul style="list-style-type: none">• Bireylerin yaşam beklentileri (Y) ile alkol tüketimleri(X) arasında bir ilişki var mıdır?• Öğrencilerin öz yeterlilikleri ile (Y) İngilizce konuşma becerileri ilişkili midir(X)?

- Korelasyon: İki değişken arasında ilişkinin ölçümü
- Korelasyon katsayısı: İki değişken arasında ilişkinin ölçümünü veren bir katsayı
- Pearson momentler çarpımı korelasyon katsayısı(r): En yaygın olarak kullanılan korelasyon katsayısı (Howel, 2013, p.185)
- Belirtme katsayısı(r^2) Korelasyon katsayısının karesidir. Verideki değişkenliğin yüzde kaçının X ve Y arasındaki ilişki ile açıklanabileceğini gösterir.
- Saçılım grafikleri: İki değişken arasındaki ilişkinin gösterilmesinde kullanılan grafikler. Her bir bireyin/katılımcının iki boyutlu bir düzlemde yerini gösterir.

Person	Family Income (in \$1000)	Student's Average Grade
A	31	72
B	38	86
C	42	81
D	44	78
E	49	85
F	56	80
G	58	91
H	65	89
I	70	94
J	90	83
K	92	90
L	106	97
M	135	89
N	174	95

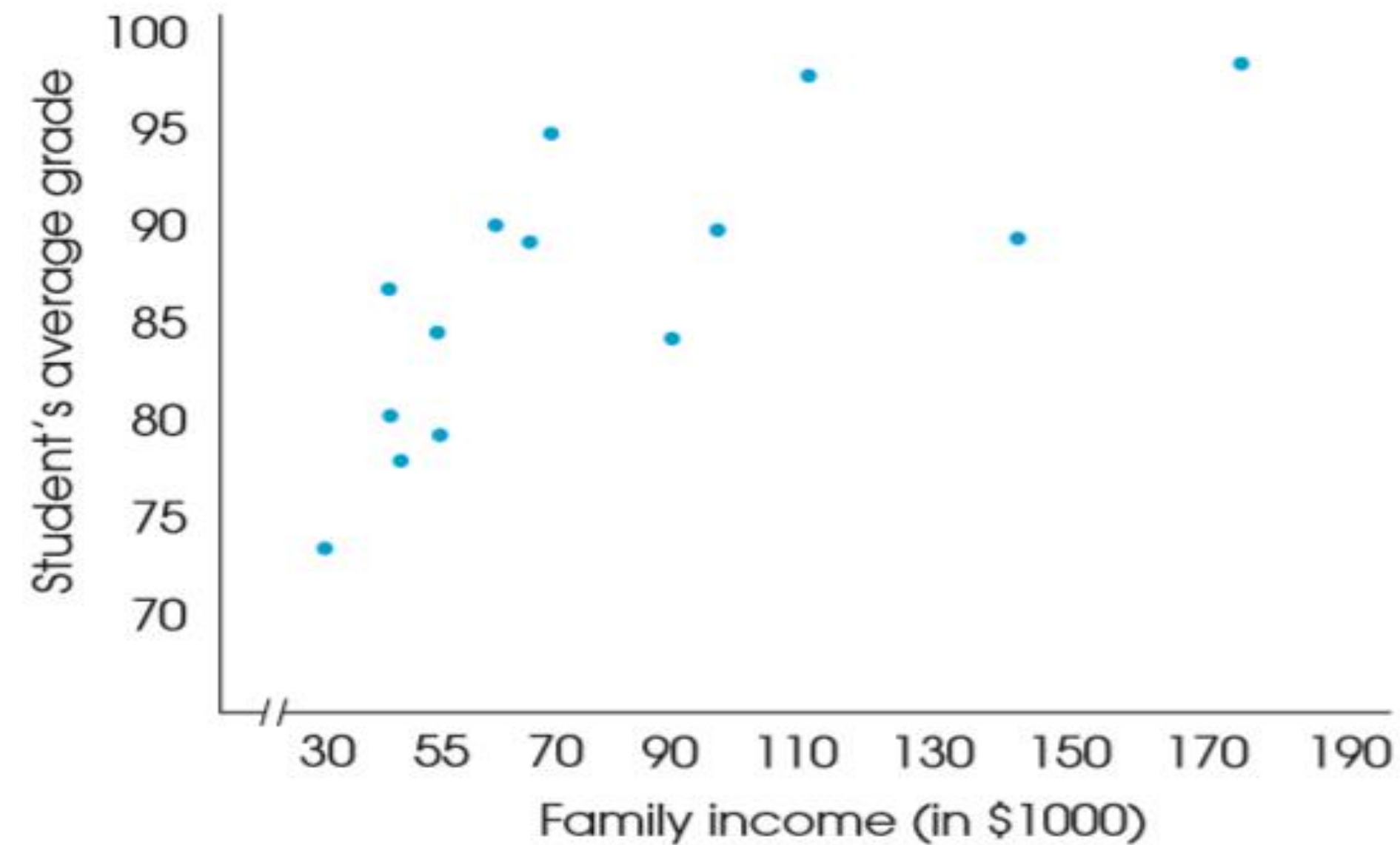


FIGURE 14.1

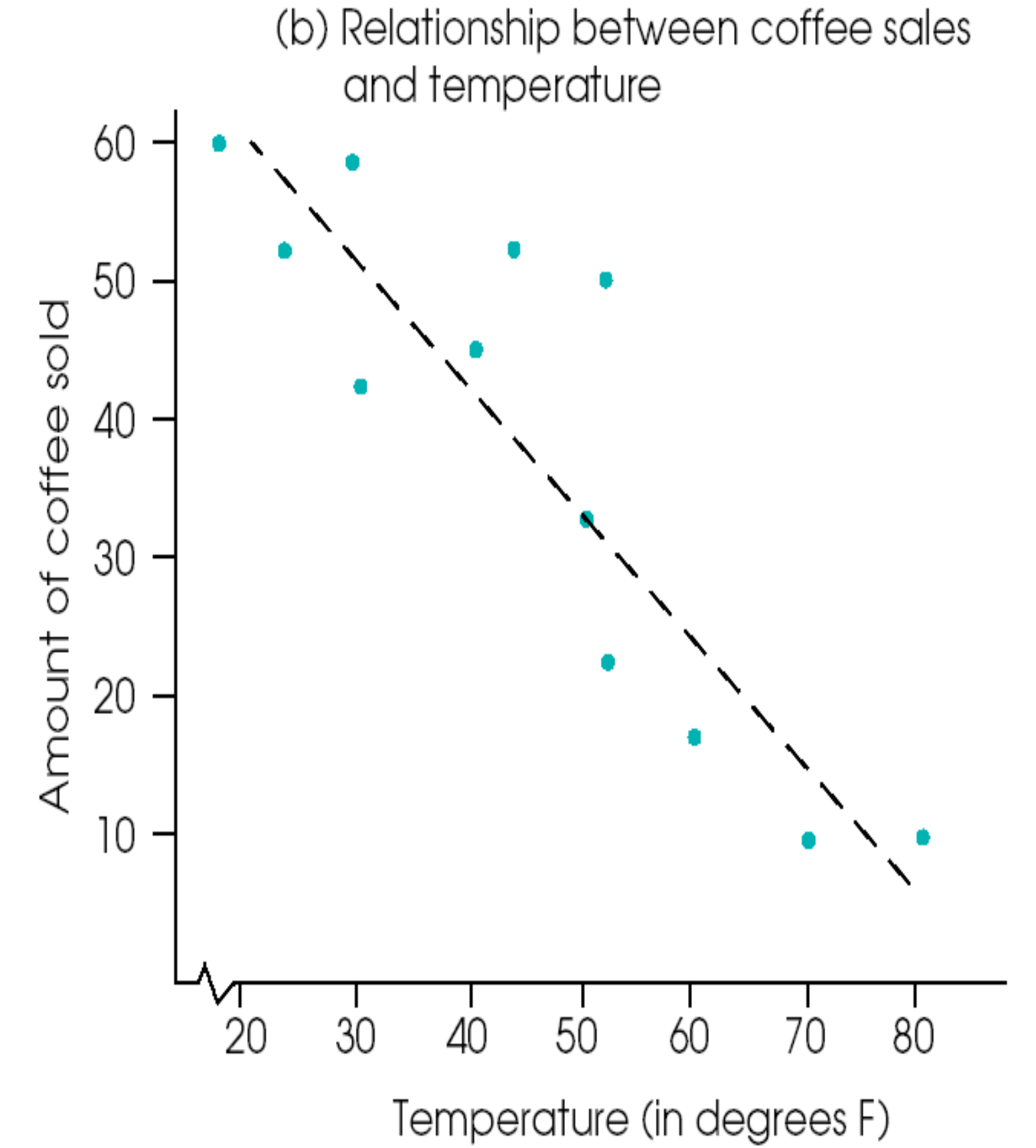
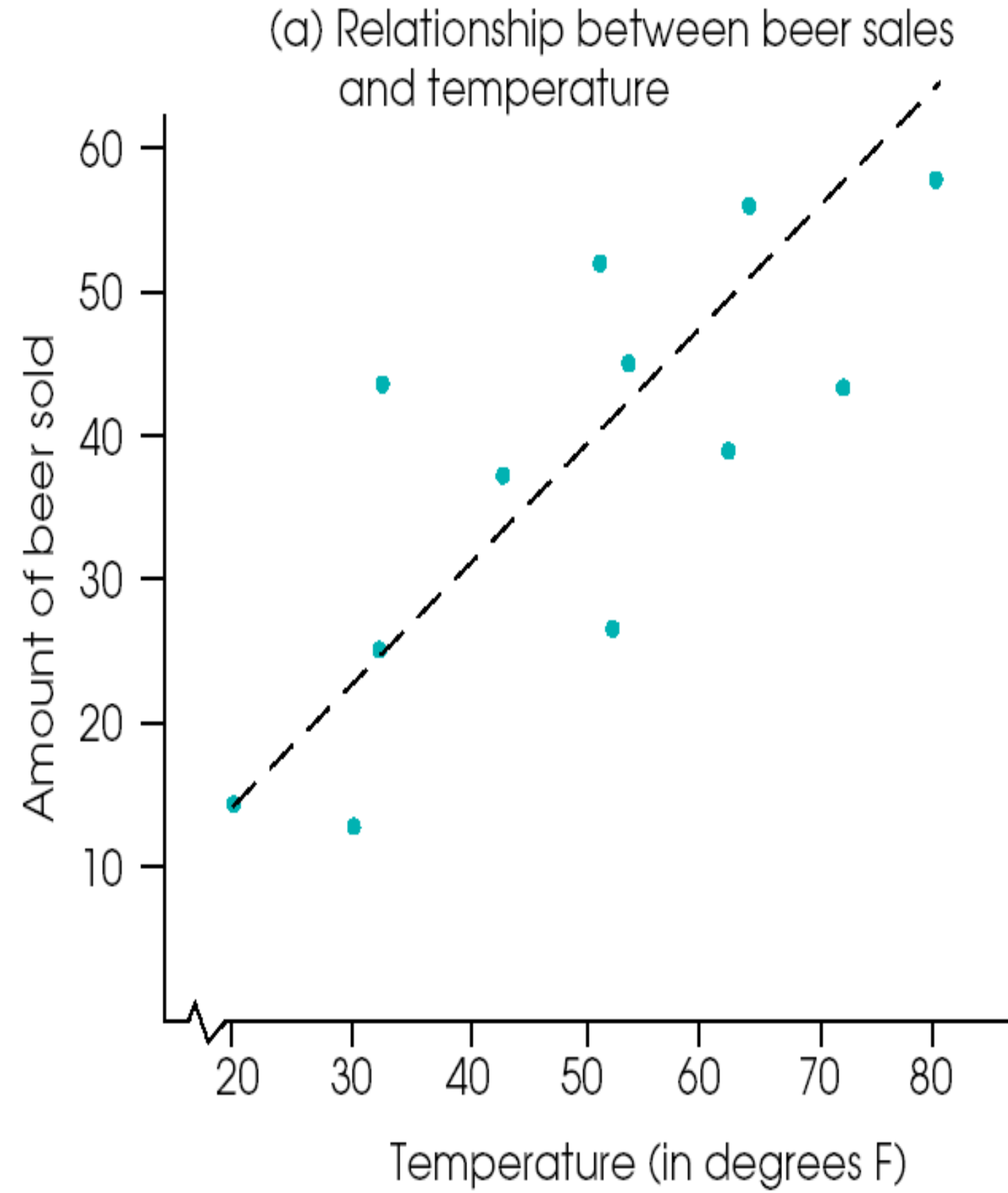
Correlational data showing the relationship between family income (X) and student grades (Y) for a sample of $n = 14$ high school students. The scores are listed in order from lowest to highest family income and are shown in a scatter plot.

İlişkilerin Özellikleri: Yön, biçim, güç

1. İlişkinin yönü

Pozitif korelasyon:

- İki değişken aynı yönde değişme eğiliminde ise
- Bir değişkenin değeri artarken diğerinin değeri de artıyorsa



Negatif korelasyon:

- İki değişken zıt yönde değişme eğiliminde ise
- Bir değişkenin değeri artarken diğerinin değeri

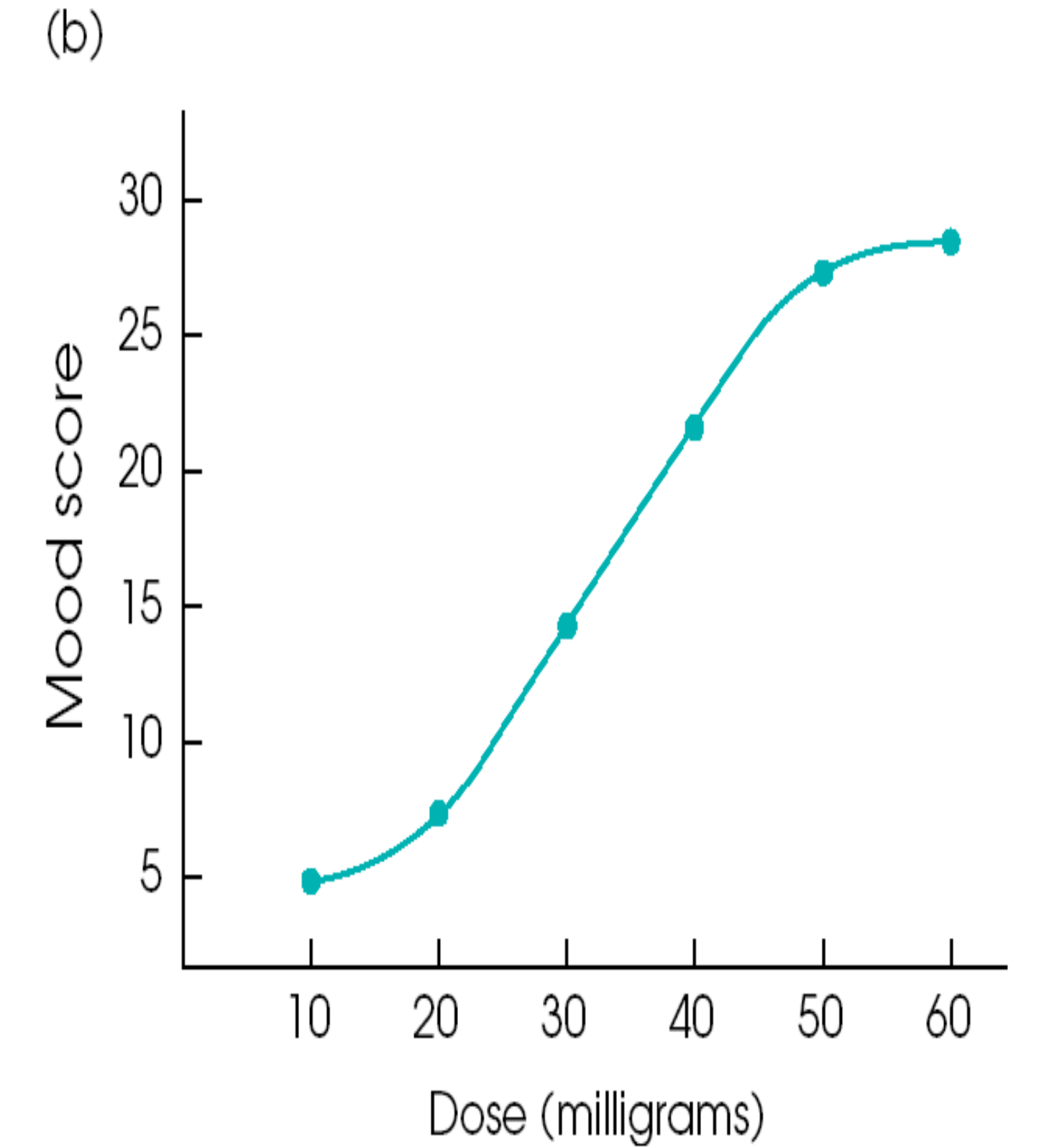
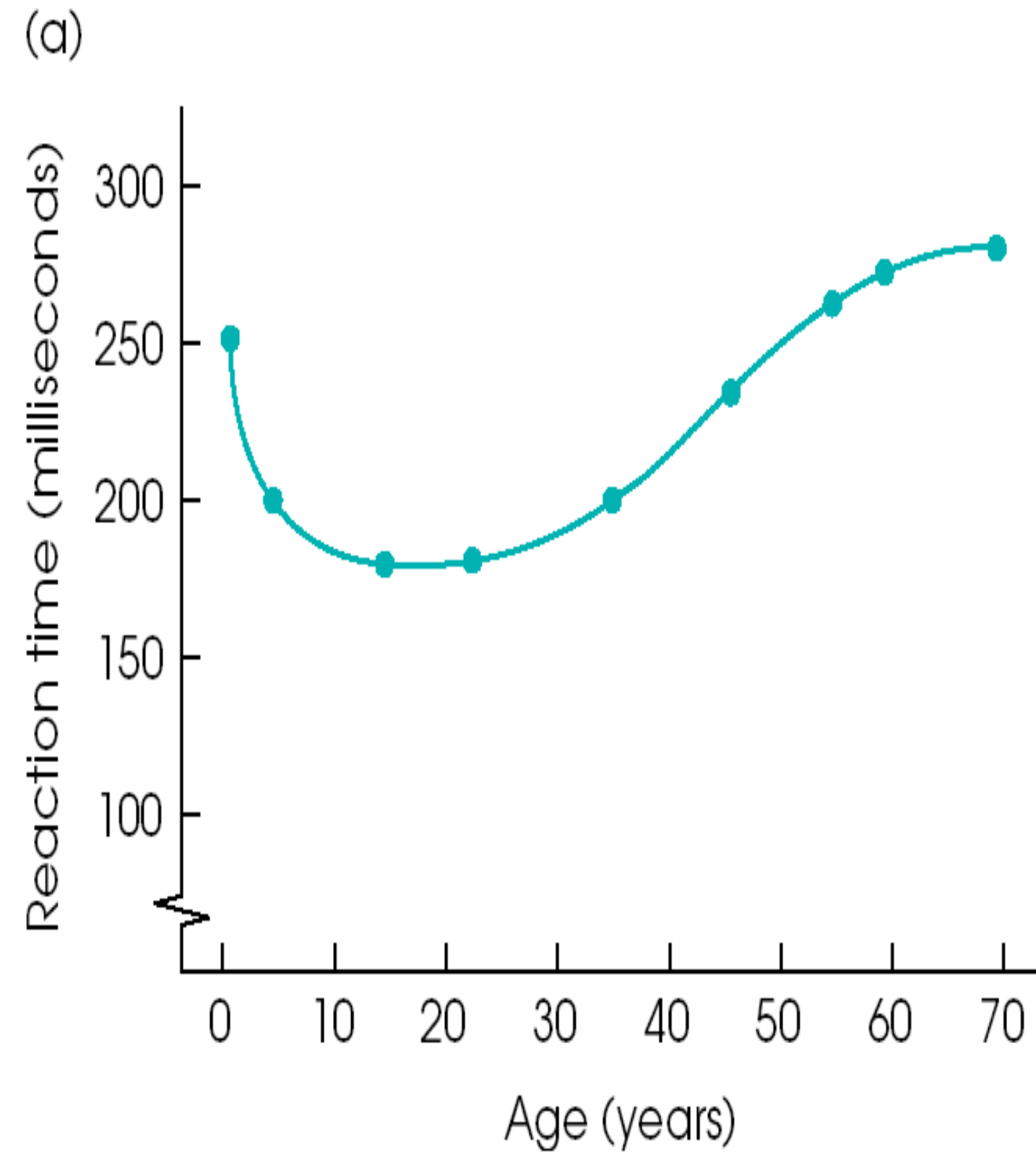
2. İlişkinin Biçimi/formu

Doğrusal İlişki:

- İlişki doğrusal bir biçimde ise
- Saçılım grafinde noktalar bir doğru etrafında toplanma eğiliminde ise,

Doğrusal olmayan ilişki:

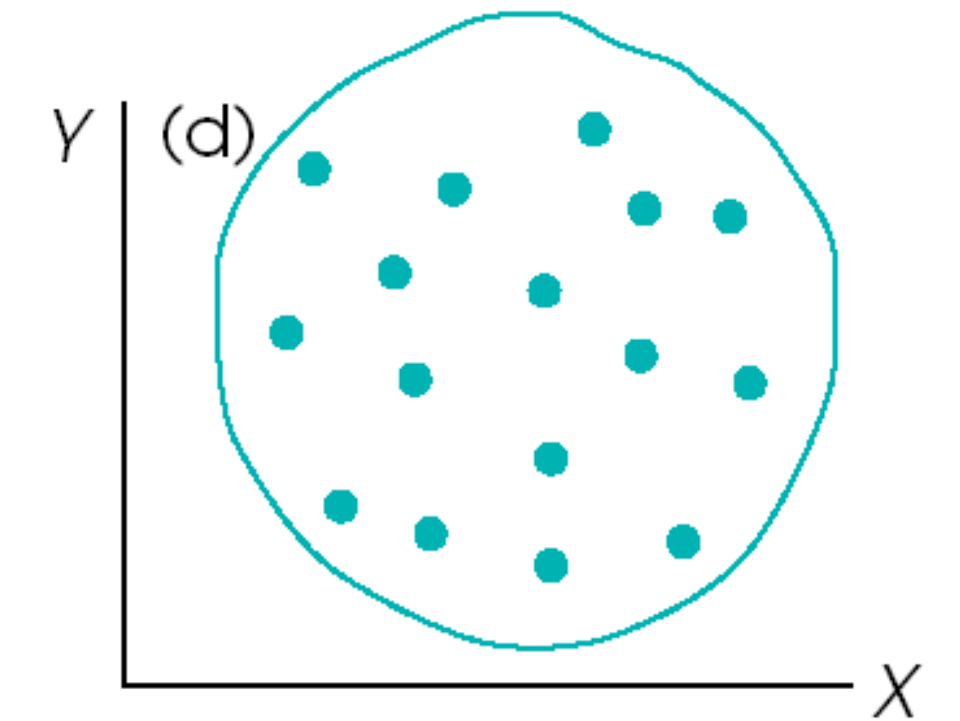
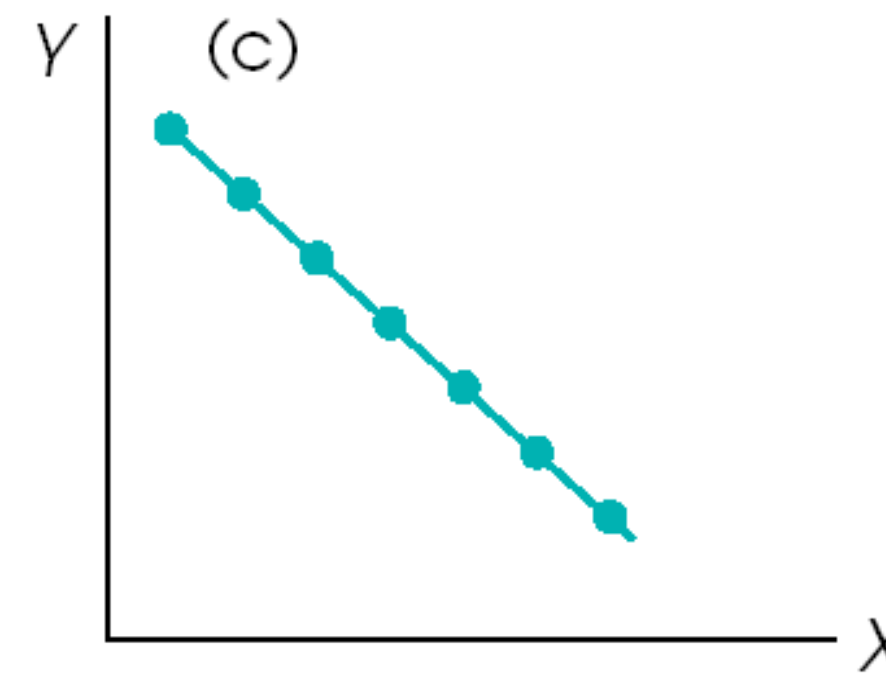
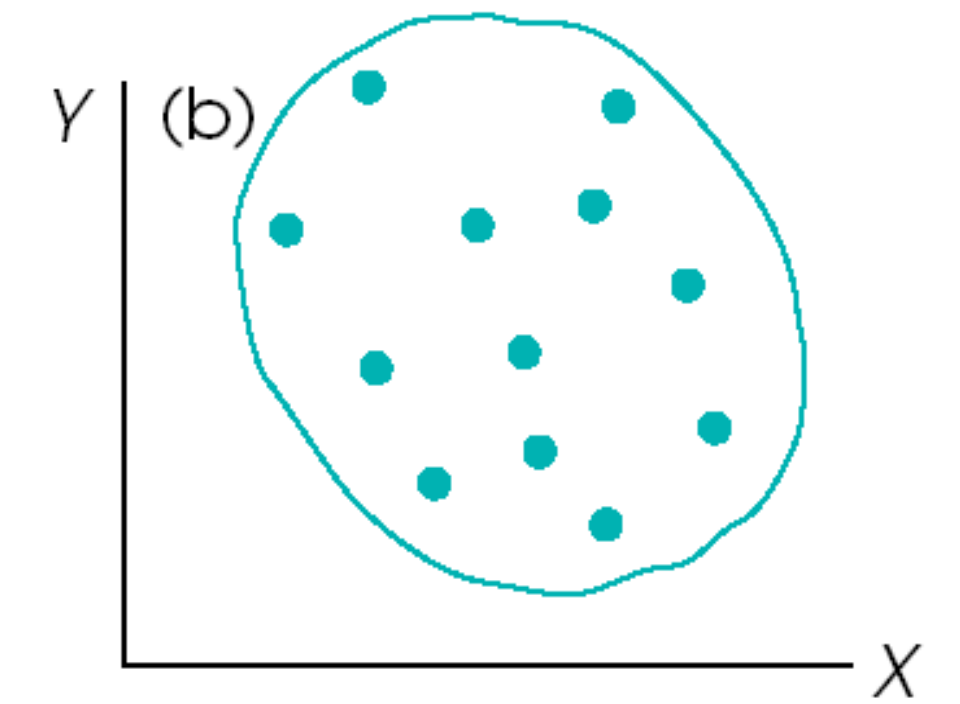
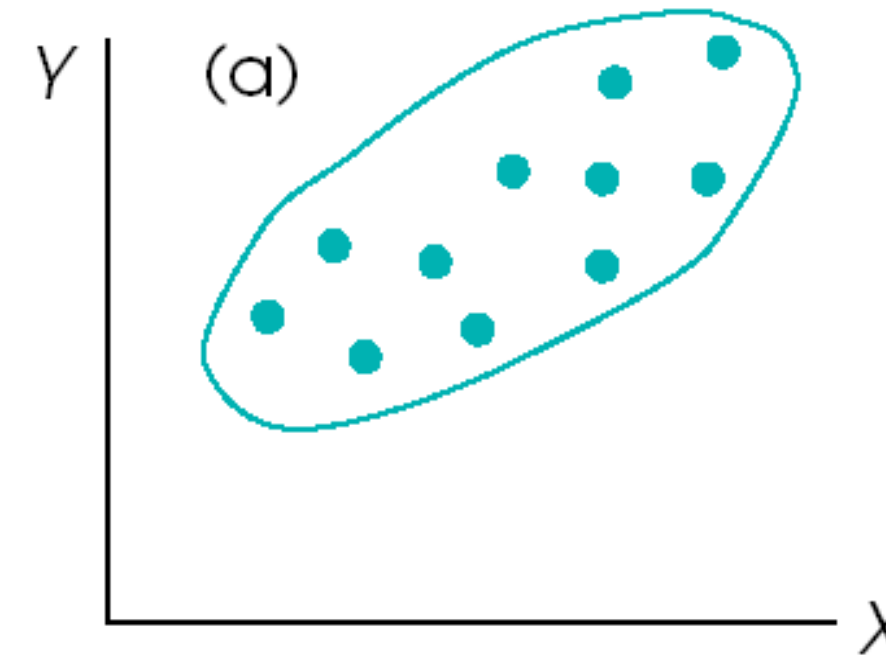
- Doğrusal ilişki tanımına uymayan tüm ilişki formları



3. İlişkinin Gücü

Mükemmel ilişki: Doğrusal bir ilişki için noktaların düz bir doğru oluşturacak şekilde toplanmasıdır (c grafiği) Relationships are usually not perfect.

- İlişkiler çoğu zaman mükemmel değildir.
- Her ne kadar X değişkeni artarken Y değişkeni de artma eğiliminde olsa (veya azalma) bu artış (veya azalış) her zaman aynı miktarda olmaz.
 - a: güçlü doğrusal pozitif ilişki
 - b: zayıf doğrusal negatif ilişki
 - c: mükemmel doğrusal negatif ilişki
 - d: ilişki yok



Korelasyon katsayısının özellikleri

- Korelasyon katsayısı -1 ile 1 arasında değerler alır.
- Katsayının işareti ilişkinin yönünü gösterir (Negatif veya pozitif)
- Katsayının büyüklüğü ilişkinin gücünü gösterir
- 1 mükemmel pozitif ilişkiyi, -1 mükemmel negatif ilişkiyi, 0 ise ilişkinin olmadığını gösterir.
- 1 ve -1'e yakın korelasyon katsayısı değerleri yüksek ilişkiyi gösterir. 0 a yakın değerler zayıf ilişkiyi gösterir.

Pearson Korelasyon katsayısı

- En yaygın kullanılan korelasyon katsayısıdır.
- r ile gösterilir.
- Varyansa dayalıdır.

Varyans: Puanların ortalama puana olan ortalama uzaklığı

Kovaryans: İki değişkenin birlikte değişkenliklerine ilişkin bir ölçüm veren istatistik.

$$r = \frac{X \text{ ve } Y' \text{nin birlikte değişme derecesi}}{X \text{ ve } Y' \text{nin ayrı ayrı değişme derecesi}}$$

$$\text{Örneklem varyansı} = s^2 = \frac{\text{Kareler toplamı}(KT)}{N - 1} = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}$$

$$\text{Örneklem Kovaryansı} = COV_{XY} = \frac{\text{Kareler Çarpımı}(KÇ)}{N - 1} = \frac{\sum_{i=1}^N (X_i - \bar{X}) * (Y_i - \bar{Y})}{N - 1}$$

$$r = \frac{KÇ}{\sqrt{KT_X KT_Y}}$$

Örnek: MEB sınıf büyüklüğü ile öğrencilerin başarıları arasında bir ilişki olup olmadığını incelemek istiyor. Bu amaçlar farklı sınıflarda öğrenim gören 15 öğrenciyi seçerek bu öğrencilerin öğrenim gördükleri sınıf mevcutlarını ve öğrencilerin sınav puanlarını kaydediyor.

Araştırma sorusu: Sınıf mevcudu ile başarı arasında bir ilişki var mıdır?

Öğrenci numarası	Sınıf Mevcudu (X)	Puan (Y)
1	19	25
2	18	23
3	12	45
4	14	40
5	7	50
6	15	22
7	19	35
8	13	32
9	17	30
10	16	31
11	15	26
12	14	37
13	20	16
14	16	28
15	19	38

Hipotez testi

- Korelasyon katsayısını genellikle örneklem verisi ile elde ederiz.
- Örneklemeler geldikleri evrenlerle birebir aynı değildir. Örneklem istatistiği ile evren parametresi arasında her zaman bir fark vardır (Örnekleme hatası)
- Buradaki temel sorumuz evrende korelasyonun var olup olmadığıdır.
- Evren için korelasyon katsayısı ρ (rho) sembolü ile gösterilir.

$H_0: \rho = 0$ (evrende korelasyon yoktur/sıfırdır.)

$H_1: \rho \neq 0$ (evrende gerçek bir korelasyon vardır.)

- Eğer korelasyonun yönü ile ilgili bir beklenti var ise hipotezler yönlü de kurulabilir.

$H_0: \rho \geq 0$ (Evrende korelasyon negatif değildir)

$H_1: \rho < 0$ (evrende negatif bir korelasyon vardır.)

Karar kuralının belirlenmesi

- Burada da t testi kullanıyoruz
- Alfa düzeyi belirlenir: 0.05
- Serbestlik derecesi: $n-2$ $15-2= 13$
- T tablosundan kritik değerler okunur: yönsüz hipotez için 2.160 ve -2.160.

Test istatistiğinin hesaplanması

$$t = \frac{\text{Örneklem istatistiği} - \text{Evren parametresi}}{\text{Standart hata}}$$

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

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

Bizim örneğimizde:

$$SE = \sqrt{\frac{1 - (-0.7)^2}{15 - 2}} = \sqrt{\frac{1 - 0.49}{13}} = 0.19$$

$$t = \frac{-0.7}{0.19} = -3.684$$

Karar: Hesaplanan t istatistiği kritik olan -2.160'dan küçük olduğu için red bölgesindedir ve H0 reddedilir.

Yorum: Sınıf mevcudu ile öğrenci başarısı arasındaki korelasyon -0.70 olarak hesaplanmıştır. Bu korelasyon anlamlı bulunmuştur
 $n=15$, $p<0.05$

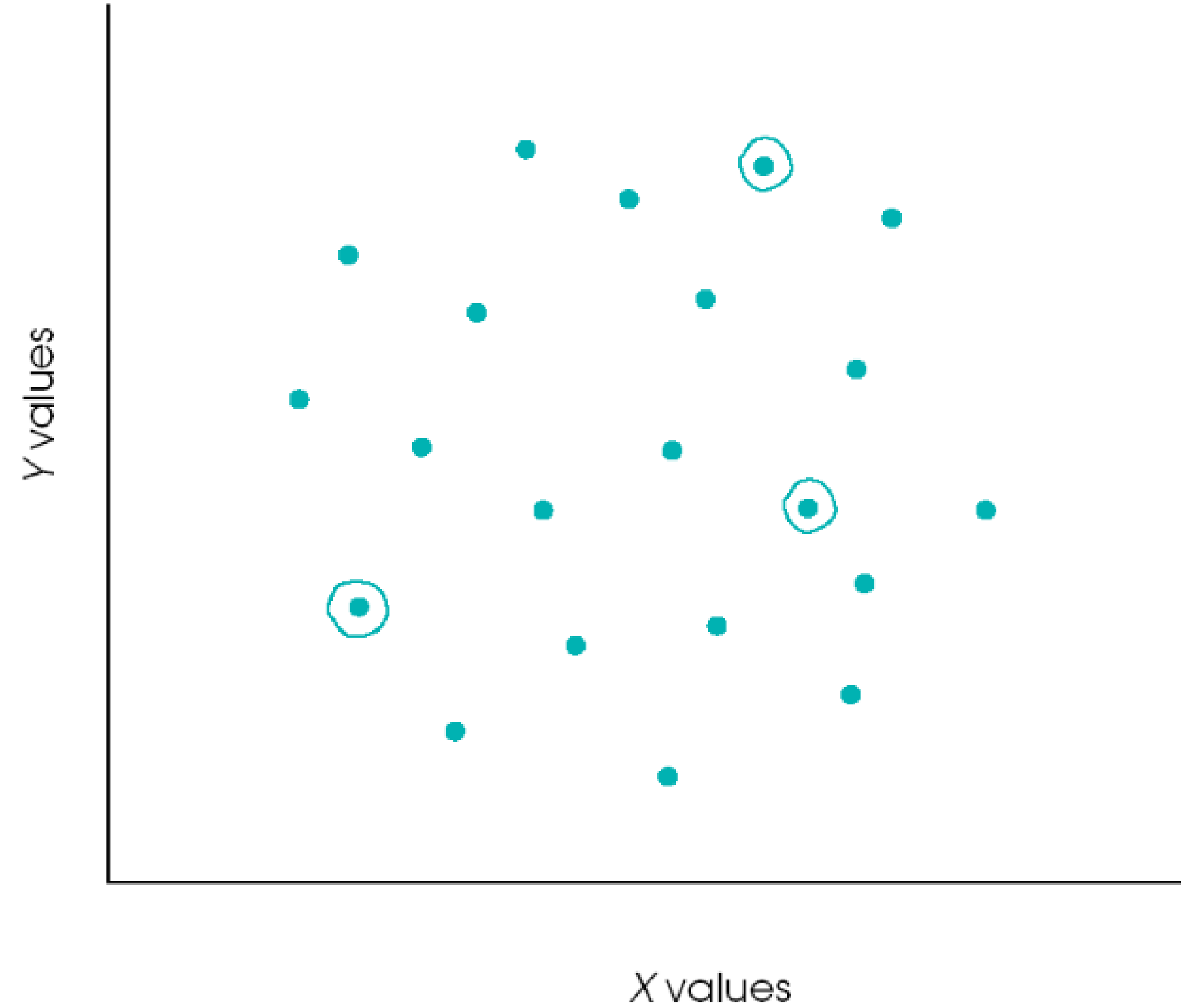
- Korelasyon yalnızca iki değişken arasında ilişki olduğunu söyler.
- İlişkinin nedenselliğine ilişkin bir bilgi **vermez**.
- **Neden sonuç ilişkisi yalnızca deneysel desenlerle yapılmış araştırmalar sonucunda elde edilebilir.**

Korelasyonın kullanıldığı yerler:

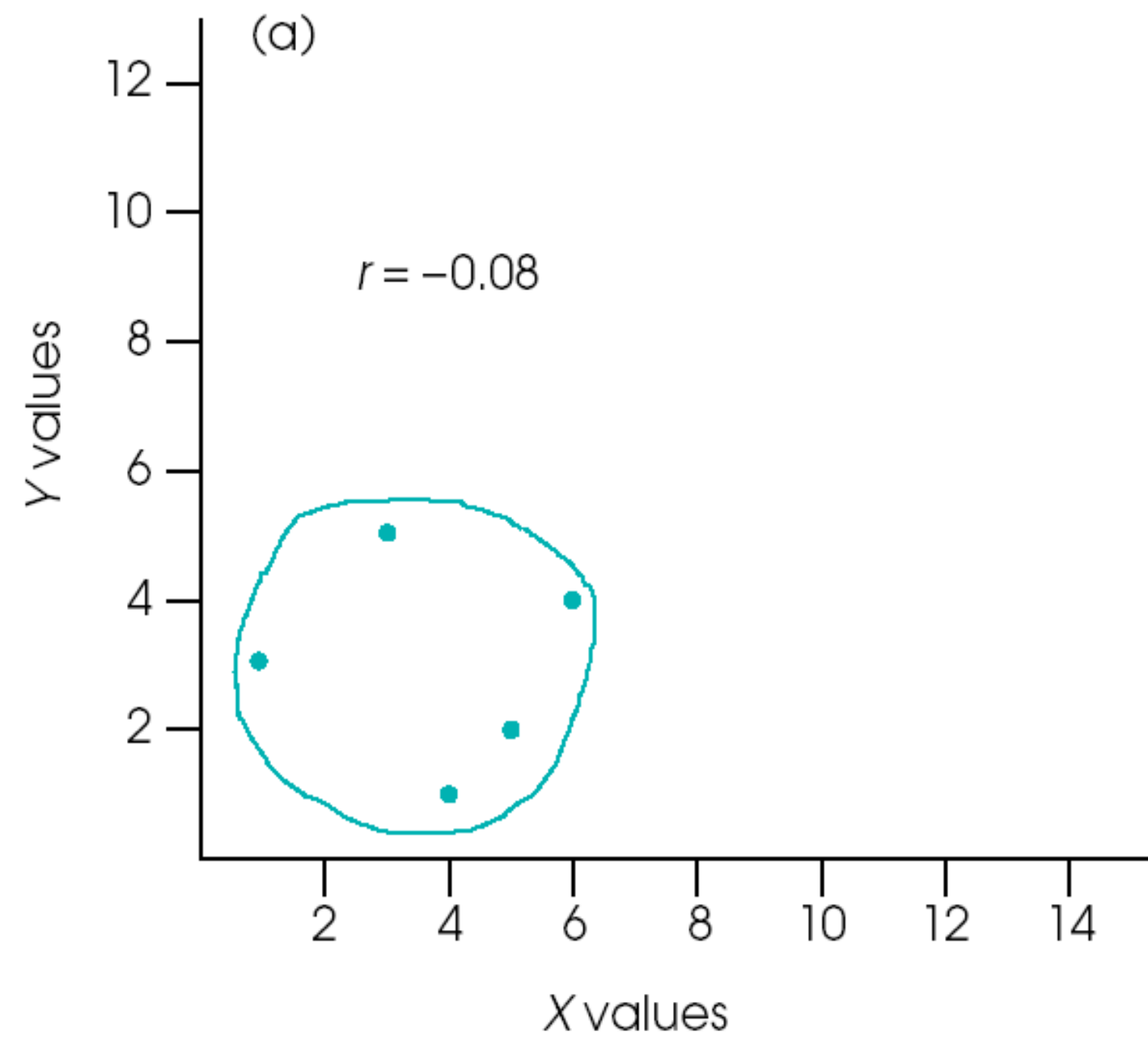
- Teori test etmek
- Kestirim yapmak
 - Eğer iki değişken arasında sistematik bir ilişki olduğunu biliyorsak bu ilişkiyi kullanarak kestirim yapabiliriz (regresyon)
- Geçerlilik
 - “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests”
 - Yeni geliştirilmiş bir testin geçerliliği yeni test ile daha önce aynı/benzer bir amaçla geliştirilmiş bir testin korelasyonu ile incelenebilir.
- Güvenirlik
 - Test puanlarının tutarlılığı
 - İki ölçüm arasındaki korelasyon güvenirliliğin bir ölçüsünü verir. (Cronbach alfa da korelasyona dayalı bir yöntemdir.)

Korelasyonu etkileyen faktörler

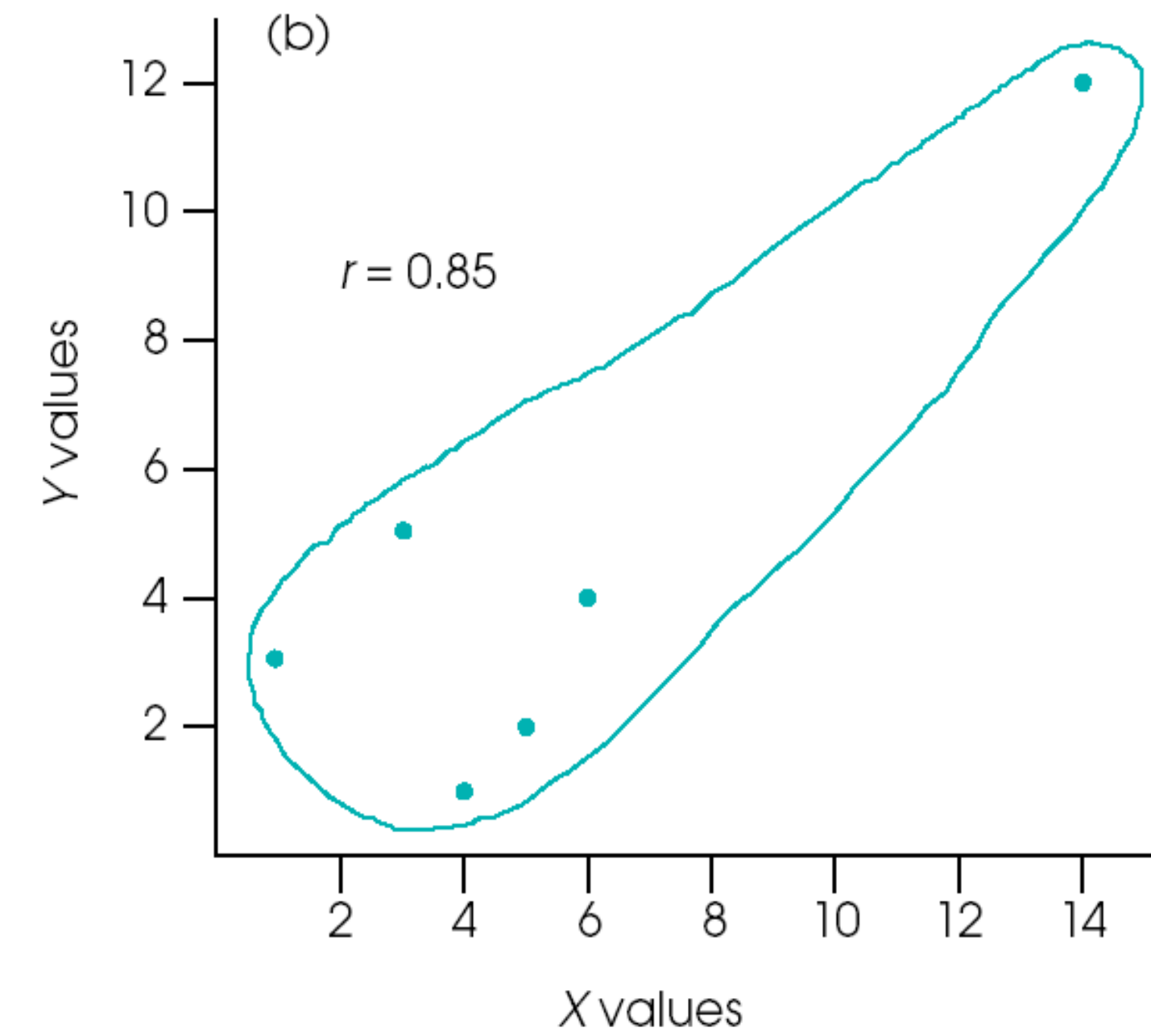
1. Örneklem Büyüklüğü



2. Uç değerler

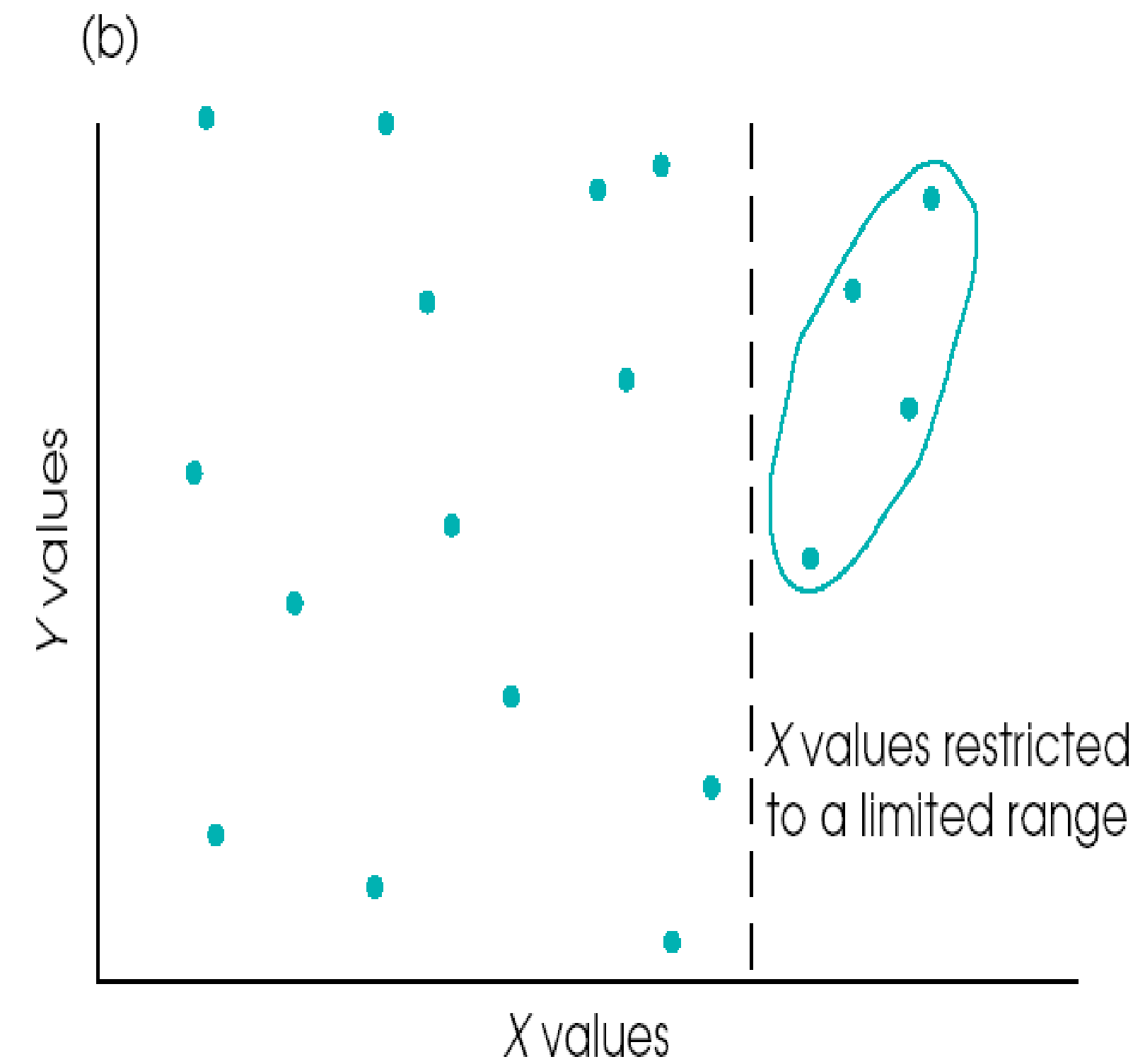
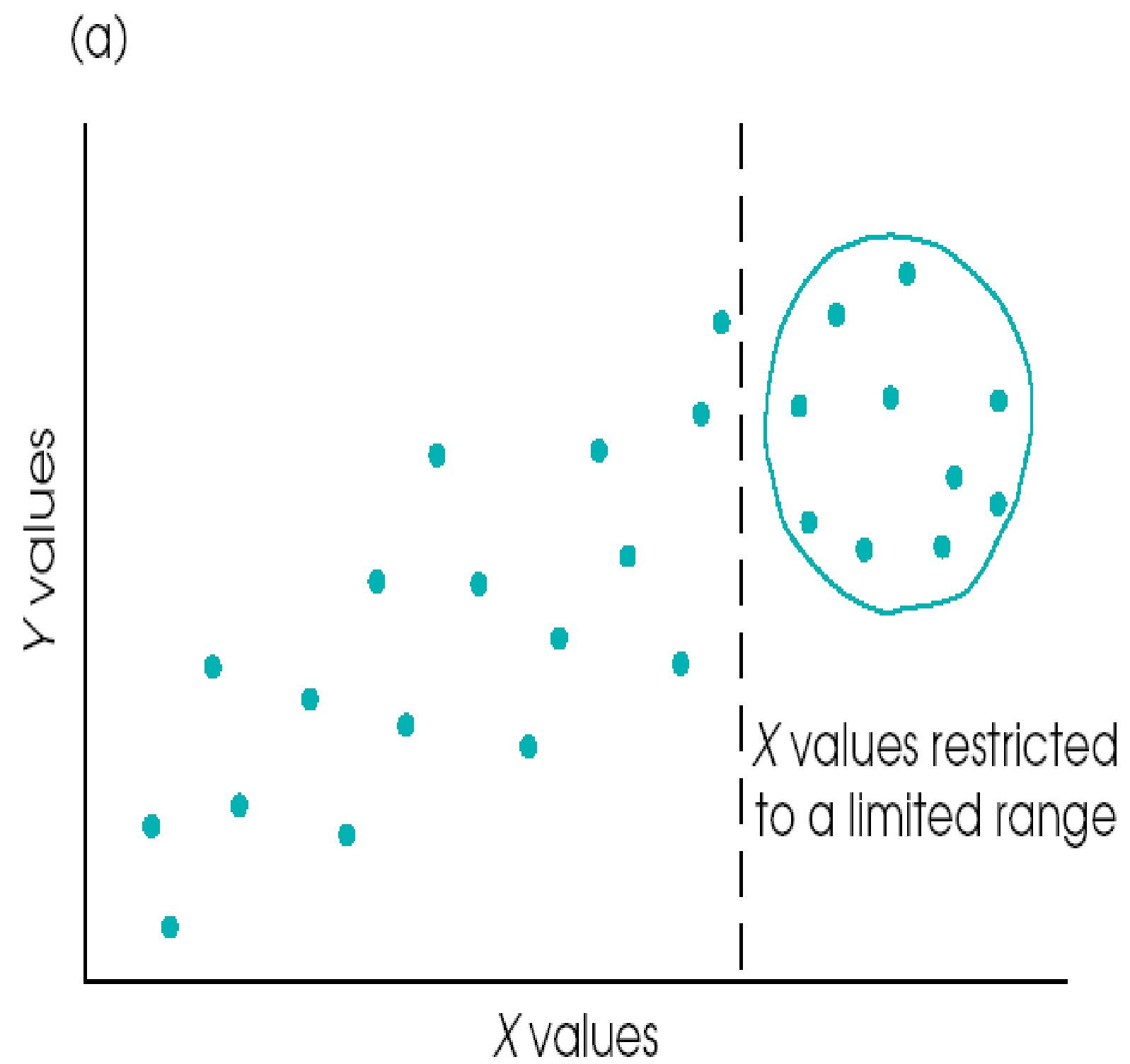


Original Data		
SUBJECT	X	Y
A	1	3
B	3	5
C	6	4
D	4	1
E	5	2

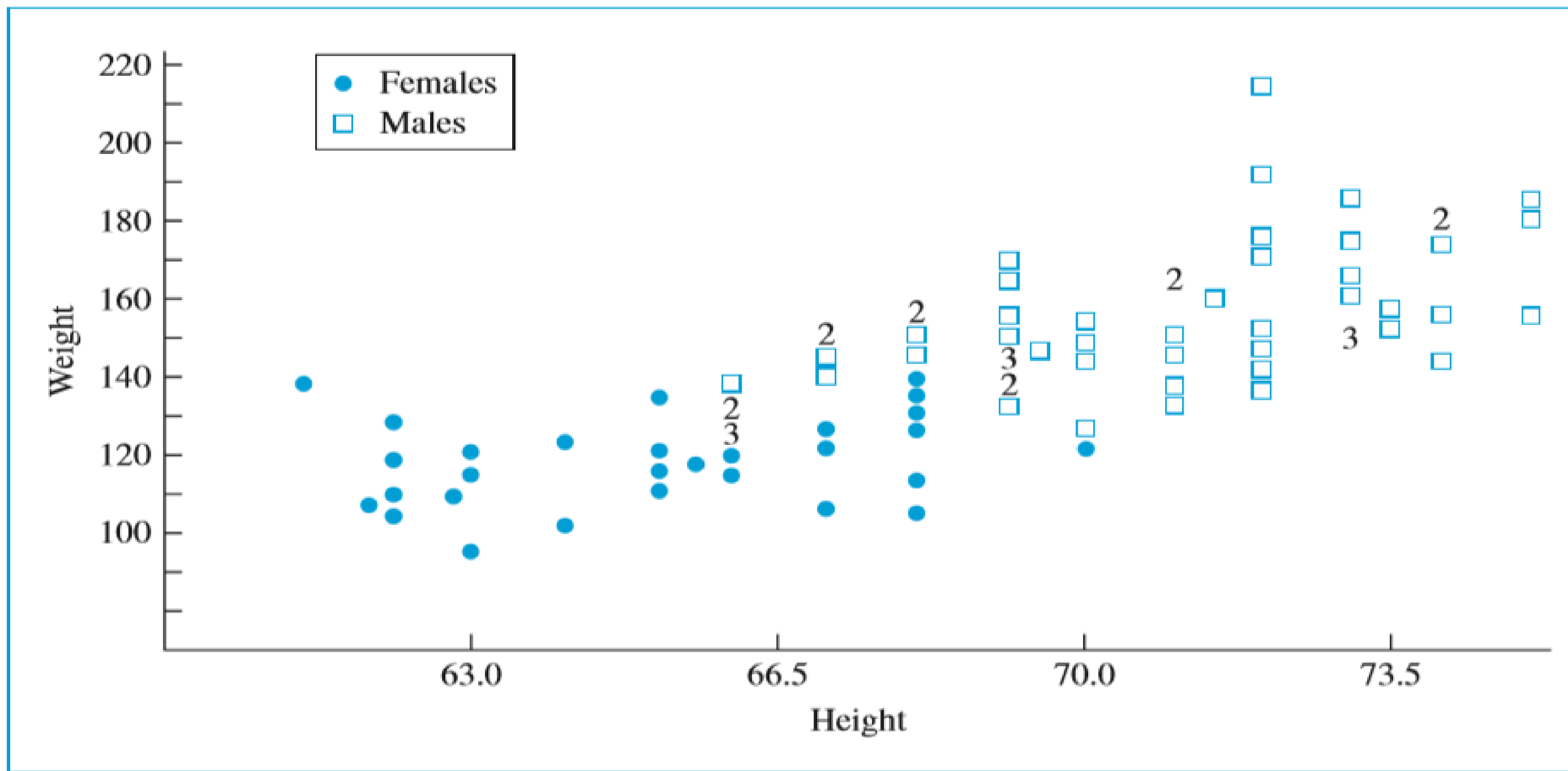


Data with Outlier Included		
SUBJECT	X	Y
A	1	3
B	3	5
C	6	4
D	4	1
E	5	2
F	14	12

3. Ranj Daralması



4. Heterojen alt gruplar



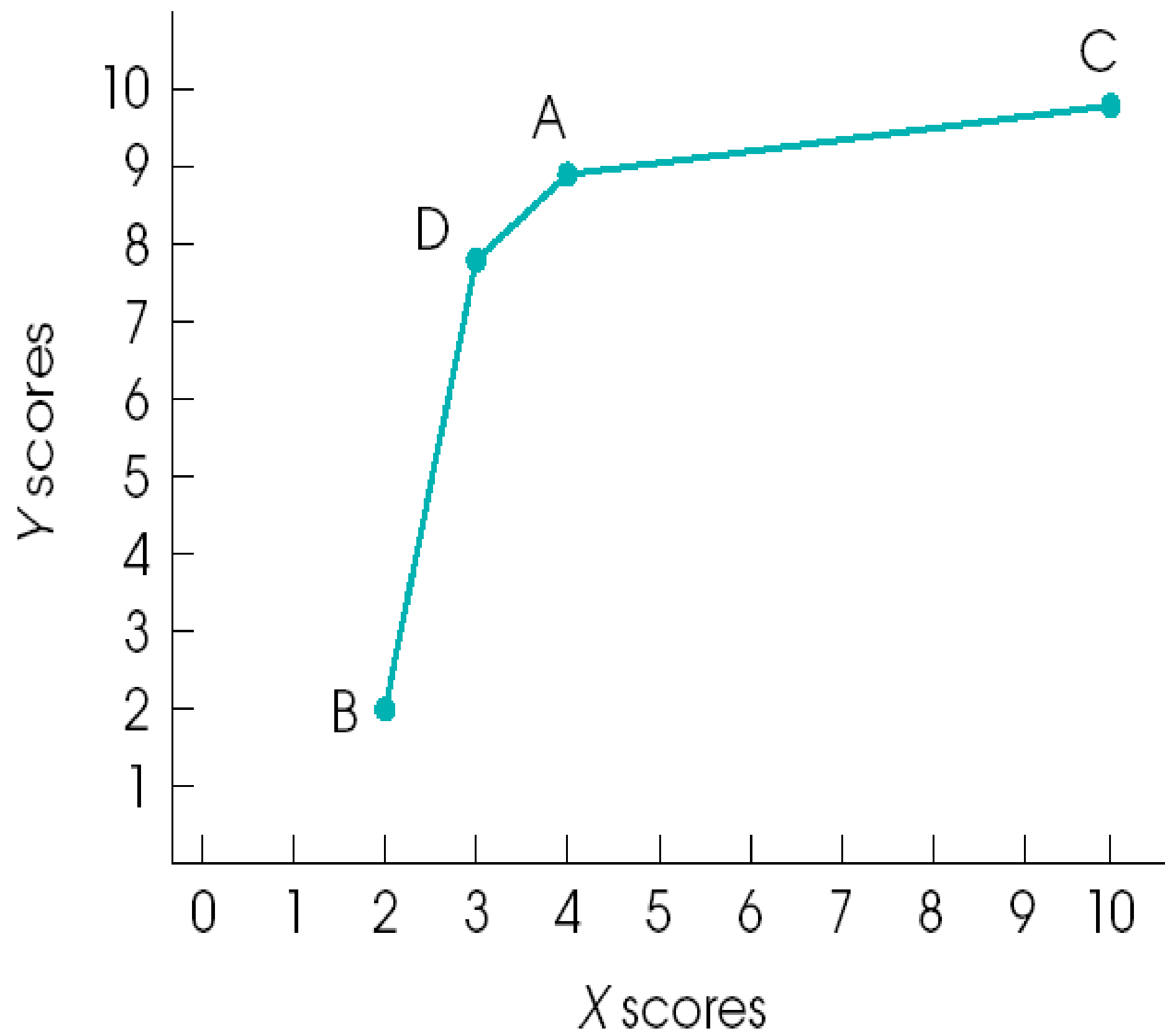
Pearson korelasyonuna alternatif yöntemler

- Pearson korelasyonu değişkenler eşit aralık veya eşit oran ölçeğinde olduğunda kullanılır.
- Doğrusal bir ilişkinin ölçümüdür.

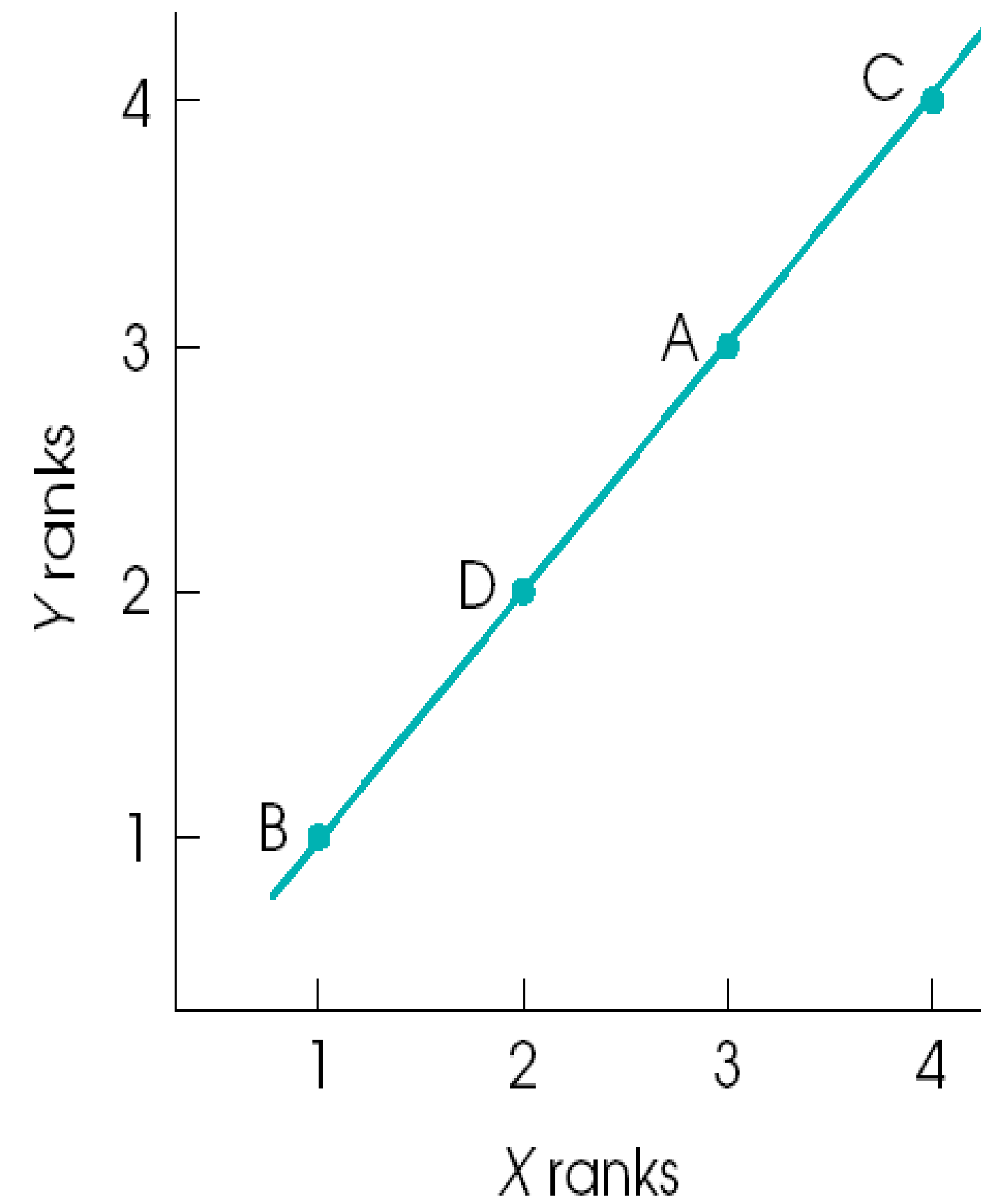
Spearman Korelasyon katsayısı

- Pearson formülü sıralama düzeyinde (sıralar) veriyle kullanıldığında elde edilen korelasyon katsayısı Spearman dır.
- Sperman korelasyonu iki değişken arasındaki tutarlılığın (consistency) bir ölçümünü verir.
- Eğer iki değişken tutarlı ise (consistent) değişkenlerin sıra numaraları arasında doğrusal bir ilişki vardır.
- Spearman korelasyonu :
 - Değişkenler sıralama ölçeğinde olduğunda veya
 - Değişkenler arasındaki ilişki doğrusal olmadığında kullanılır.

(a) Scores



(b) Ranks



- Pearson formülünde ham puanlar yerine sıra numaraları kullanılarak elde edilir.
- r_s ile gösterilir
- Puanlar eşit olduğunda sıra numaralarının ortalaması alınır.

Puanlar		Sıralar		Sıra sapmaları		Sıralar için kareler		Kareler çarpımı
(X)	(Y)	X	Y	$X-M_X$	$Y-M_Y$	$(X-M_X)^2$	$(Y-M_Y)^2$	$(X-M_X)^* (Y-M_Y)$
19	25	5	2	2	-1	4	1	-2
18	23	4	1	1	-2	1	4	-2
12	45	2	4	-1	1	1	1	-1
14	40	3	3	0	0	0	0	0
7	50	1	5	-2	2	4	4	-4
Ortalama Sıra		3	3			10	10	-9

Nokta çift serili korelasyon katsayısı (Point biserial)

- İki değişkenden biri numerik diğeri iki kategorili olduğunda kullanılır.
- İki değişken arasındaki ilişkinin gücüne ilişkin bir ölçüm verir.
- Nokta çift serili korelasyon katsayısını hesaplayabilmek için:
 - İkili puanlanan değişken 0 ve1 olarak kodlanır.
 - Standard pearson formülü uygulanır.

Phi Katsayısı

- Değişkenlerin ikisi de iki kategorili olduğunda kullanılır.
- Phi katsayısı hesaplanırken
 - İkili puanlanan değişkenler 0 ve 1 olarak kodlanır.
 - Standard pearson formülü uygulanır.