

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 manipulasyan yapmayız, kontrol etmeye çalışmayız.
- Değişkenleri bağımlı veya bağımsız olarak sınıflandırmayız.

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.

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

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

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

$$r = \frac{KÇ}{\sqrt{KT_X KT_Y}}$$

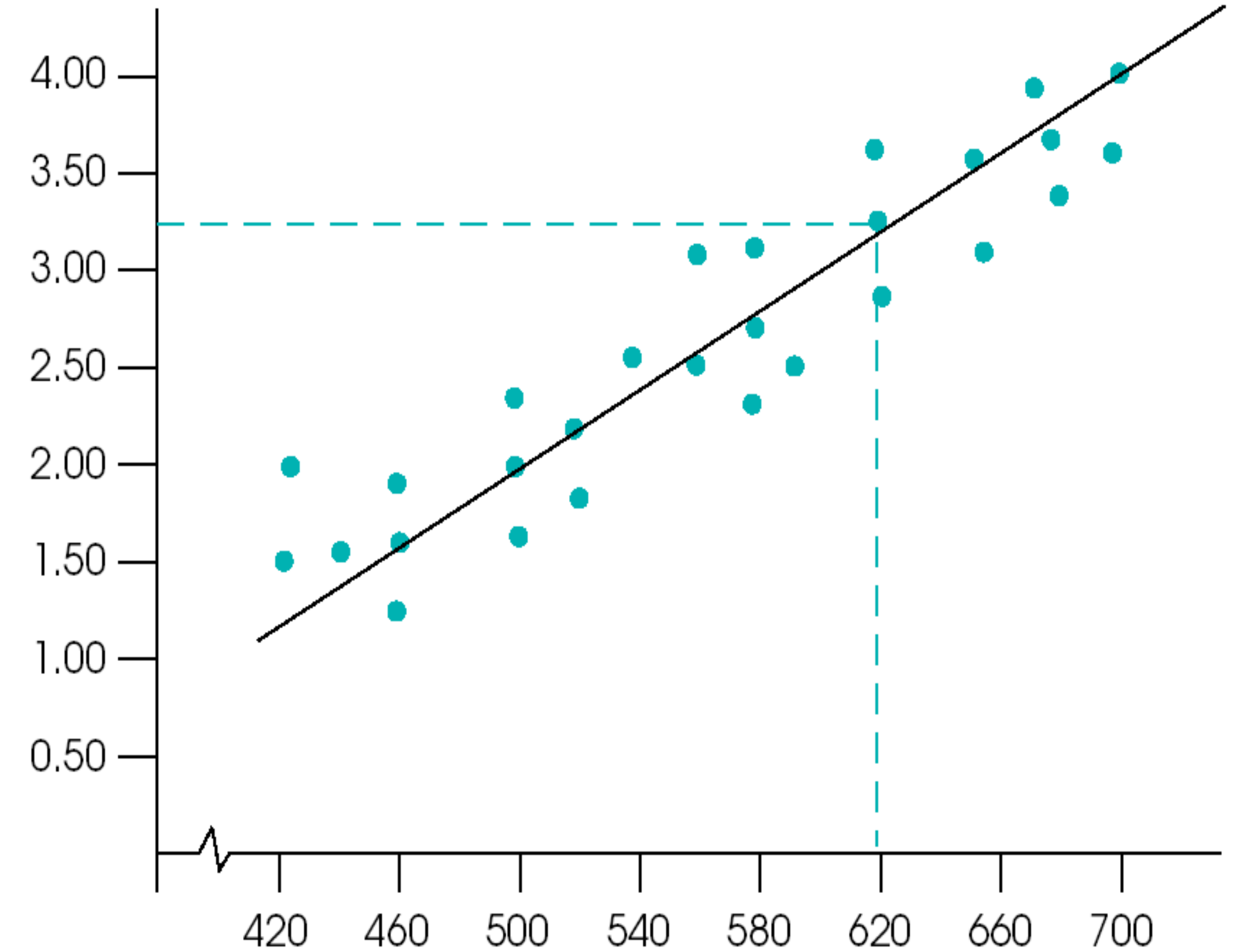
Örnek: MEB sınıf büyüklüğü ile öğrencilerin başarıları arasında bir ilişki olup olmadığını incelemek istiyor. Bu amaçlar farklı sınıflarda öğrenim gören 15 öğrenciyi seçerek bu öğrencilerin öğrenim gördükleri sınıf mevcutlarını ve öğrencilerin sınav puanlarını kaydediyor.

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

Basit Doğrusal Regresyon

- Korelasyonu kullanarak kestirim yapabiliriz.
- İki değişken arasındaki ilişki saçılım grafiğiyle gösterilebilir.
- Bu iki değişken arasındaki ilişkiyi bir doğru ile de gösterebiliriz.
- Bu doğruya regresyon doğrusu denir.
- Regresyon analizlerinde temel amaç veriye en iyi uyum sağlayan regresyon doğrusunu elde etmektir. (best fitting line)



- Yapmaya çalıştığımız bir değişkendeki değişimin diğerini nasıl etkileyeceğini açıklayan bir denklem oluşturmaktır.
- Diğer bir ifadeyle regresyon doğrusunun denklemini elde etmeye çalışıyoruz.
- Bir değişkeni kullanarak diğer değişkenin değerini tahmin etmek için bir denklem oluşturmaya çalıştığımız durumlarda aslında regresyon analizi yapıyoruz.
- Regresyon analizinde kestirmeye çalıştığımız değişken (Y değişkeni) bağımlı değişken, kestirici değişken (X değişkeni) bağımsız değişken olarak tanımlanır.

Doğru Denklemleri

- X ve Y değişkenleri arasındaki doğru denklemi:

$$Y = b X + a$$

- b eğim katsayısıdır. Eğim X değişkenindeki bir birimlik değişimin Y değişkenini ne kadar değiştirdiğini gösterir.
- a kesişim katsayısıdır. X değişkeni 0 değerini aldığı anda Y değişkeninin değeridir.

Örn: İstanbul'da taksi tarifesi 5 lira ile başlar ve kat edilen her bir kilometre için 3 lira yazmaktadır. İstanbul'da taksi tarifesi için bir denklem yazabiliriz:

$$Y = 3X + 5$$

- Bir doğru üzerindeki iki noktayı biliyorsak doğrunun denklemini yazabiliriz.

$$Y = bX + a$$

$$4 = b \cdot 1 + a \quad a = 4 - b$$

X	Y
1	4
5	16

$$16 = 5b + a$$

$$16 = 5b + 4 - b$$

$$4b = 12 \quad b = 3 \quad a = 1$$

Regresyon

- Bir veriye en iyi uyum sağlayan doğruyu bulmak için kullanılan tekniğe regresyon denir.
- Bir veri setinin merkezden geçen pek çok doğru çizilebilir.
- Burada problem veriye en iyi uyum sağlayan doğruyu bulabilmektir.

En iyi uyum sağlayan doğru ne demektir?

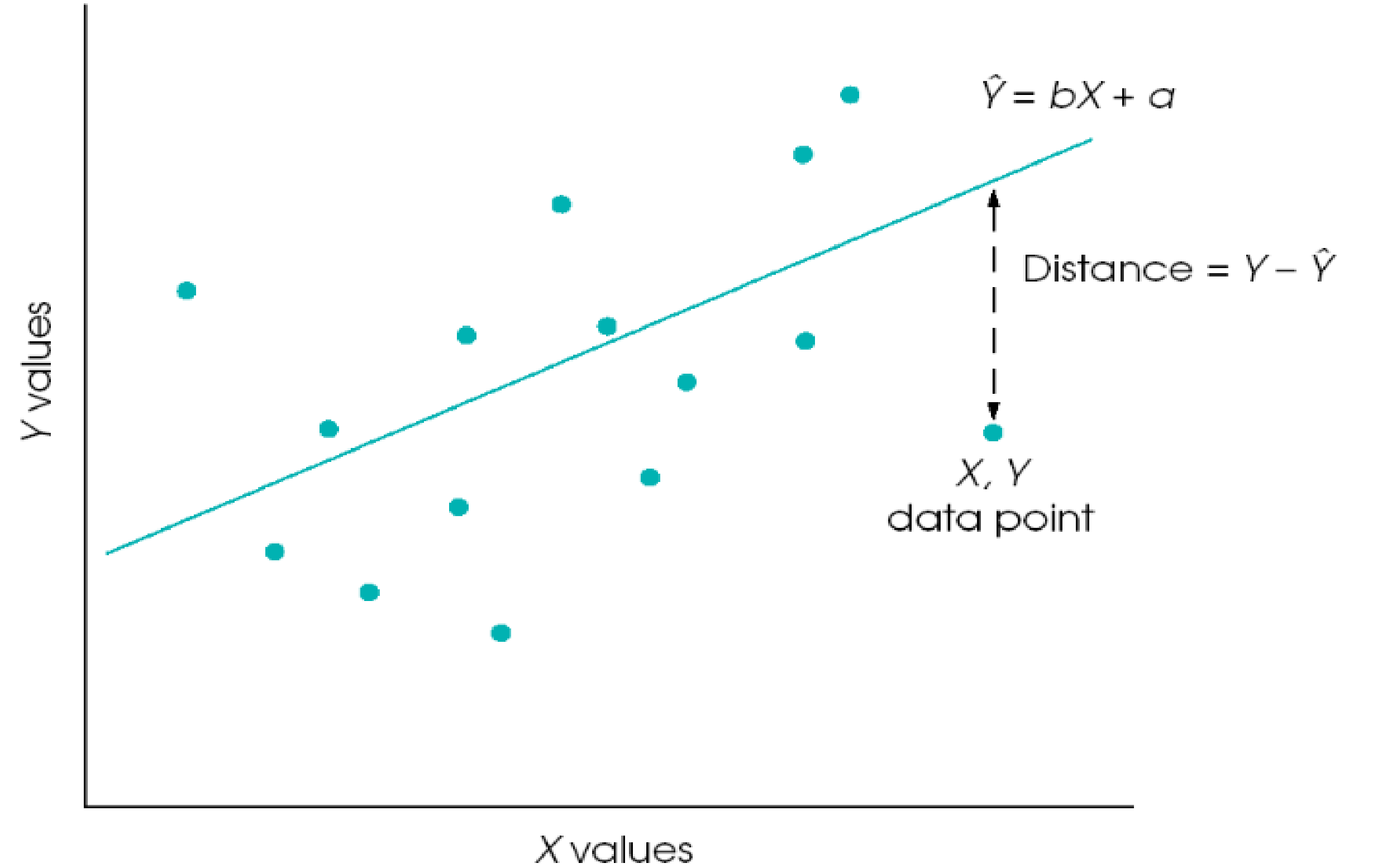
- Matematiksel olarak bir nokta ile bir doğru arasındaki mesafeyi ölçebiliriz.
- Her bir X değeri için iki tane Y değerimiz olur. Birisi veriden gelen Y değeri, diğeri denklemden hesapladığımız \hat{Y} değeridir.
- En iyi uyum gösteren doğru bu iki Y değeri arasındaki farkın en küçük olduğu doğrudur.

Mesafe (Hata) = $Y - \hat{Y}$

- $Y - \hat{Y}$ değeri negatif değerler de alabilir.
- Bu nedenle farkın karelerini alırız.

Toplam hata kareleri = $\sum (Y - \hat{Y})^2$

- Toplam hata karelerini minimize eden a ve b değerlerini bulmak bazı hesaplamalar gerektirir.



$$\hat{Y} = bX + a$$

$$\bullet b = \frac{KÇ}{KT_x}$$

$$\bullet a = My - bMx$$

- KÇ: Kareler Çarpımı = $\sum_{i=1}^N (X_i - \bar{X}) * (Y_i - \bar{Y})$
- KT_x: X in Kareler Toplamı = $\sum_{i=1}^N (X_i - \bar{X})^2$
- r= X ve Y arasındaki korelasyon katsayısı
- M_y= Y nin ortalaması
- M_x= X in ortalaması

Burada a ve b katsayıları artık regresyon katsayıları olarak da isimlendirilir.

Örnek: Bir önceki örneğimizi kullanırsak

Araştırma sorusu: Sınıf mevcudu öğrencilerin başarılarını yordayabilir mi?

$$\text{Başarı} = \text{SınıfMevcudu} * b + a$$

Standartlaştırılmış regresyon katsayıları

- Bazen araştırmacılar regresyon denklemini hesaplamadan önce X ve Y puanlarını standart z puanlarına dönüştürürler.
- Standart puanlar kullanılarak elde edilen regresyon katsayılarına standartlaştırılmış regresyon katsayıları denir.
- Regresyon denklemini basitleştirmek için kullanılır.

$$\widehat{z}_y = (beta) * z_x$$

- SPSS her iki katsayıyı da raporlar

Kestirimin standart hatası

- Regresyon bize bir değişkene ilişkin kestirim yapma olanağı verir.
- Ancak analizin kendisi bize kestirimin ne kadar iyi olduğu hakkında bilgi vermez.
- Kestirimin hassasiyetine ilişkin bilgiyi bize kestirimin standart hatası verir.
- Standart hata bize gözlenen Y değerleri ile kestirilen Y değerleri arasındaki ortalama mesafeyi verir.

$$SH = \sqrt{\frac{\sum (Y - \hat{Y})^2}{N - 2}}$$

Regresyon analizinde hipotez testleri

- Korelasyon analizinde olduğu gibi regresyon denklemlerini de 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 regresyon katsayılarının anlamlı olup olmadığıdır.

- Evrende regresyon denklemimiz:

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

- Burada β evrende eğimdir, α evrende kesişim katsayısıdır, ε hata terimidir. ε rastgele hatadır ve evrende ortalaması sıfırdır.
- Biz hipotez testlerimizde β ve α katsayılarının evrende sıfırdan farklı olup olmadığı üzerine kurarız.
- Bizim için önemli olan β katsayısının evrende sıfırdan farklı olup olmadığıdır.