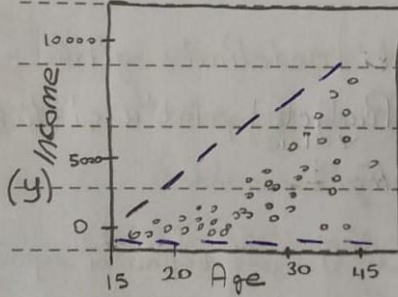


Heteroscedasticity and Homoscedasticity

(Değişen Varyans)

(Sabit Varyans)

Değişen Varyans:



=> Bağımlı ve Bağımsız değişken arasındaki Heteroscedasticity durumu

* Biz Regresyon analizinde Error'lar bağlamında "Değişen Varyans" tan söz ediyoruz. Spesifik olarak, heteroskedisite; error'ların gerçek değerler aralığına yayılmasındaki sistematik değişikliklerdir.

* Homoscedasticity, sabit varyansın eşit ve scatter'ın homojen bir yapıya sahip olma durumu iken Heteroscedasticity yani Değişen Varyans bunun tam tersidir.

Peki Bunlara Nasıl Bakmalıyız?

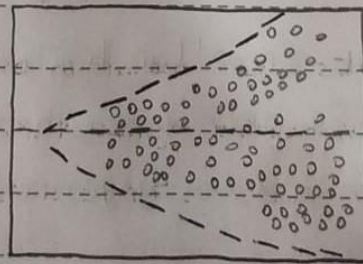
1) * Heteroskedastisite yani Değişen varyans, Errorlarımızın büyüklüğünün bağımsız bir değişkenin değerlerine göre farklılık gösterdiği yani homojen bir yapıda olmadığı durumda görülür.

→ Errorların, bağımsız değişkenin low veya high değerleriyle nasıl değiştiğine bakılır.

2) * Errorların varyansı sabit olmalıdır, ki modelimiz güvenilir olsun. Bu ne demek? Error - Predicted plot'u çizdirip buradaki varyans değişimini gözlemeliyiz.

ÖR

Residual (error)



Predicted Value
(fitted value)

→ Sabit Varyans formuna uymuyor.
(Değişen varyans sorunu var)

⇒ Error'ların varyansı bir değişkene bağlı olarak değişebilir bu yüzden

1. adımıdaki işlem yapılarak problemlili değişken bulunur.

Değişen Varyans Sorunu Durumunda

1) Plot Bağımlı ve Bağımsız Değişken ve Dağılıma Bak

2) Error - Target Plot'u Çizdir

3) Error - Bağımsız değişken plot'u çizdir ve sorunun kaynağını bul.

* Nasıl aksiyonlar alınabilir?

Örneğin;

Elimizde Family Balance bağımsız değişkeni var ve harcama ile ilişkisine bakıyoruz. Bu durumda çok fazla larger ve lower değer olacak. Bu durumu düzeltmek için

Target(harcama)'ın harcama oranı olarak revize edilmesi;

Family Balance değişkeninin "cap" lenmesi; Dönüşüm yapılması gibi birçok metod denenebilir.

Değişen Varyansa son bir örnek:

↳ Imagine, we use family income to predict luxury spending (Y)

Beklenen üzere; income ile harcama arasında positive güçlü

bir ilişki vardır. Error's are very small for low values of

family income (fakir ailelerin hiçbiri lüks harcama

fakat; great variation of error's for zengin

families (bazı zenginler lüks alır; bazı almaz çünkü)

Bu durum "Heteroskedasticity" sorunudur.