

Normallik Testleri

İstatistiksel yöntemlerin birçoğu normallik varsayımına dayanır. Anova, Regresyon, t testi gibi testler verilerin normal dağılıma sahip olduğu durumlarda kullanılmaktadır. Bu nedenle de analize başlamadan önce veri setinin dağılımının normal olup olmadığının test edilmesi gerekir.

Eğer veri setinin dağılımı normal değilse veri seti aşağıda gösterilen alternatif yaklaşımlardan bir tanesi kullanılarak analiz edilmelidir.

Bu alternatif yaklaşımlar

- Parametrik Olmayan Teknikler
- Dayanıklı (Robust) Teknikler
- Dönüşümler

olarak belirtilebilir.

Literatüre girmiş birçok normallik varsayımı testi vardır. Bunlar şöyle sıralanabilir:

- Anderson – Darling Testi
- Shapiro – Wilk Testi
- Kolmogorov – Smirnov Tek Ornek Testi
- Pearson Ki Kare Uygunluk Testi
- Cramer- Von Mises Testi
- Shapiro – Francia Testi

Bu testler birbirinin türevi olarak şekilde çoğaltılmıştır. Yani bu listeyi daha da genişletmek mümkündür. Yapılan çalışmalarda ise Shapiro-Wilk testi ve Anderson-Darling testi diğerlerine göre daha güçlü bulunmuştur. Bu sebeple kullanılan paket program ya da R ve Python gibi istatistik ve veri bilimi alanında kullanılan yazılım dillerinde farklı tiplerde testler uygulanabilmektedir. Örnek vermek gerekirse SPSS`te Shapiro – Wilk ve Kolmogorov Smirnov testleri Minitab programıysa Anderson-Darling, Ryan Joiner ve Kolmogorov Smirnov testlerini sunmaktadır.

Dersimizde faydalandığımız R programında ise kullanıcıların ekledikleri paketlerle bu metotlar sürekli artış gösterebilmektedir. Temel olarak bulunan Shapiro – Wilk testi herhangi bir paket yüklemeye gerek kalmadan “shapiro.test” komutuyla hemen uygulanmaktadır. Bu testin yanında “nortest” paketinde bulunan diğer testleri de kullanabiliriz. Bu pakette aşağıdaki normallik testleri bulunmaktadır.

- ad.test Anderson-Darling normallik testi
- cvm.test Cramer-von Mises normallik testi
- lillie.test Lilliefors (Kolmogorov-Smirnov) normallik testi
- pearson.test Pearson chi-square normallik testi
- sf.test Shapiro-Francia normallik testi

Burada test edilen hipotezler şu şekildedir.

H_0 : Verilerin geldiği dağılım ile normal dağılım arasında farklılık yoktur.

H_1 : Verilerin geldiği dağılım ile normal dağılım arasında farklılık vardır.

Bu hipotezlerin yaygın olarak yazılımı ise şu şekilde olmaktadır.

H_0 : Veriler normal dağılmıştır.