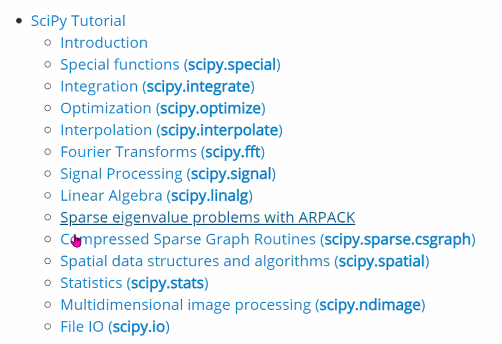
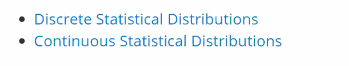
Scipy kütüphanesi. Geniş bir kütüphanedir. (Scintific python) (scipy.org sitesinden araştırılabilir)

İçeriği (sub modülleri) şöyledir:



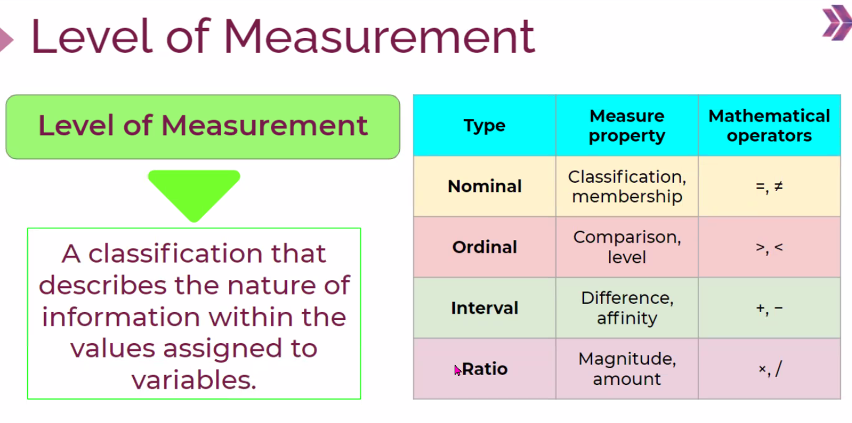
Scipy.stats modülünün içinde



bulunmaktadır.

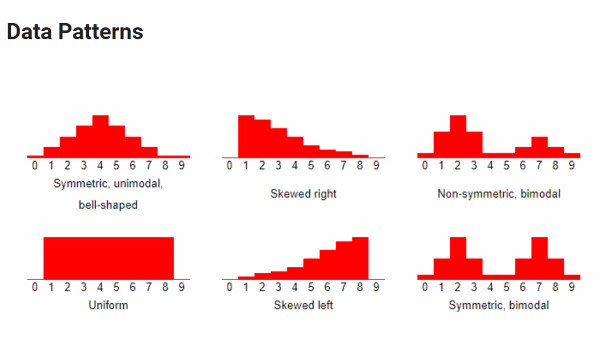
Özellikler summary statistics üzerine çalışacağız.

Bu modül, çok büyük sayıda olasılık dağılımının yanı sıra büyüyen bir istatistiksel işlev kitaplığı içerir.



Nominal hiyerarşik değil, ordinal hiyerarşiktir. Aralarındaki fark budur.

İnterval – ratio : Bu ikisine scale’de denir. ratio özellikle age gibi yaş gibi sıfırdan başlayan değerlerdir. Aralarında oran vardır. 4 yaşındaki çocuk 2 yaşındaki çocuğun iki katı yaşındadır. İnterval sıcaklık derecesi gibi aralarında oran olmayan verilerdir.



Amaç saklı paternleri ortaya çıkarmaktır. Her daha bize bir şey söyler aslında ama başta o gizlidir, bilmeyiz. Onu ortaya çıkarmak için (değişkenler arasındaki ilişkiler, tek değişkenin şekli vb.) her saklı paternden bir yorum çıkarabiliriz ve böylece söyleyecek bir şeyimiz olur.

İstatistiğin temeli , bir popülasyondan sample alarak bunun istatistik verilerinden yola çıkarak popülasyon hakkında genel bilgiler elde etmektir. Popülasyon ifade eden herşey parametre oluyor. Sample ifade eden her şey de istatistikdir.

İstatistikler hep küçük harflerle simgelenir. Popülasyon ise büyük harflerler gösterilir.

Mean üzerine yaklaşımda varyans ve standart sapma bizim kullanacağımız iki değerdir. Mean tam orta noktayı tanımlıyor yani data setimi tam olarak tamamlayamıyor. O yüzden varyans ve standart sapma ile ilave tanımlamalar yapıyoruz.

Varyans bir yayılım, bir dağılım kavramı. Standart sapma ise varyansın kareköküdür.

Madian based deyince akla Q1, Q2, Q3, Q5, IQR (Q3-Q1) kavramları gelmeli. Veriyi quartile kavramlarla açıklıyoruz.

Percentile ise veriyi 100 parçaya bölüyor. 25 % , quartile yaklaşımına göre Q1’e eşittir.

50%, Q2 (median)’a eşittir. 75% ise Q3’e eşittir.

İstatistikte bizim derdimiz hep outlier ve missing values'lar. Bu whisker plot'i (Turkler box plot diyor) yorumlamayi iyi öğrenmek lazim. Iyi anlanirsa outlier'larin tespitinin yaninda arkada yatan distribution ve skewness'i de anlayabilecegiz kedi gibi biyiklari olan bir kutucu