

R Kursu Ödev 5 Tanımlayıcı İstatistikler ve İki Grup İçin Testler

1- MASS paketindeki `Pima.tr` verisi bir grup kadın hastadan alınmış bilgilerden oluşur. Bu verisetinde `npreg` hamilelik sayısını, `glu` kan şekeri ölçümünü, `bp` kan basıncını, `skin` koldan alınan bir ölçüm değerini, `bmi` beden kitle indeksini, `ped` şeker hastalığı ile ilgili bir değeri, `age` yaşı ve `type` şeker hastası olup olmadığını gösteren değişkenlerdir. İstatistiksel testlerde aslında belirli hipotezler test edilir. Mesela şeker hastası olan ve olmayan hastalarda `bmi` değerlerinin ortalaması aynıdır gibi. Siz de bu verisetini kullanarak aşağıdaki soruları cevaplayınız.

* Bu verisetindeki her değişkene uygun tanımlayıcı istatistikleri bulunuz.

* Şeker hastası olan ve olmayan hastalarda `glu`, `bp`, `skin` ve `age` değerlerinin aynı ortalamaya sahip olup olmadığını t testi kullanarak test ediniz. (Önce varyansların homojen olup olmadığına bakınız)

* Yukarıda yaptığınız işlemi nonparametrik bir yöntem kullanarak test ediniz. Sonuçları t testi sonuçları ile karşılaştırınız.

* Şeker hastası olanlarda `bmi` değerlerinin ortalamasının şeker hastası olmayanlara göre daha yüksek olduğunu istatistiksel olarak araştırınız.

2- `iris` verisini kullanarak `virginica` türü çiçeklerde `Sepal.Length` ortalamasının `setosa` türü çiçeklere göre daha yüksek olduğunu test ediniz. (Önce dataframe üzerinde `virginica` ve `setosa` türü çiçekleri seçmek/süzme gereklidir).

3- MASS paketindeki `anorexia` verisini kullanarak CBT ve FT gruplarındaki `Prewt` ve `Postwt` değerlerini karşılaştırınız. Hangi grupta kilo alımı daha fazla olmuştur ve istatistiksel olarak anlamlı mıdır? (2.soruya benzer şekilde CBT ve FT gruplarına ait verileri seçmek gereklidir).

4- Kendinize ait bir verisetinde `t.test` ve `wilcox.test` testlerini yapınız.

Başarılar.