

<https://www.kaggle.com/datasets/fedesoriano/body-fat-prediction-dataset?resource=download>

adresinde bulunan **bodyfat** veri setinden **set.seed (öğrenci numaranız)** komutunu kullanarak **150** büyüklüğünde veriyi rasgele seçiniz.

1. Veri setinde yer alan değişkenlerden seçtiğiniz 3 tanesi için bildiğiniz tanımlayıcı istatistik yöntemleri kullanarak, sonuçları yorumlayınız. (değişken seçiminde, öğrenciler arasında kesişimin minimum sayıda gerçekleşmesi beklenmektedir.) **(10 puan)**
2. Veri setinde yer alan değişkenlerin dağılımlarını inceleyerek, 3 farklı dağılıma sahip (normal dağılıma en yakın dağılıma sahip değişken(1), sağdan çarpık dağılıma sahip değişken(2), soldan çarpık dağılıma sahip değişken(3)) değişkeni seçerek normal dağılıp dağılmadıklarını Cramer-von\_Mises, Anderson Darling, Kolmogorov Smirnov, Shapiro Wilk uyum iyiliği testleriyle hipotezleri oluşturarak test ediniz. Normallik testlerini yaparken kullandığınız p-değeri ne anlama gelmektedir? Eğer normal dağılmıyorlarsa uygun dönüşüm tekniklerini kullanarak değişkenlerin dağılımının normale yaklaşıp yaklaşmadığını kontrol ediniz. **(30 puan)**
3. Yukarıda linki verilen veri setinden çektiğiniz 150 büyüklüğündeki veri kitle olarak kabul edildiğinde, daha önce seçtiğiniz farklı dağılımlara sahip (normal dağılıma en yakın dağılıma sahip değişken(1), sağdan çarpık dağılıma sahip değişken(2), soldan çarpık dağılıma sahip değişken(3)) 3 değişken için 1000 tekrarlı, örneklem büyüklüğü 5, 10, 30 ve 50 olan örneklem çekerek Merkezi Limit Teoremini ispatlayınız. **(30 puan)**
4. Yukarıda linki verilen veri setinden çektiğiniz 150 büyüklüğündeki veri kitle olarak kabul edildiğinde, seçtiğiniz bir değişken için 30 büyüklüğünde rassal bir örneklem çekiniz.
  - %99 güven düzeyinde, kitle ortalaması için güven aralığı oluşturunuz ve yorumlayınız. Güven aralığı bulunurken hangi dağılım kullanılmaktadır. Gerekçesiyle anlatınız. **(10 puan)**
  - %90 güven düzeyinde, kitle varyansı için güven aralığı oluşturunuz ve yorumlayınız. Güven aralığı bulunurken hangi dağılım kullanılmaktadır. Gerekçesiyle anlatınız. **(10 puan)**
5. Bir ilaç şirketi, yeni üretilen bir ilacın belli bir hastalığa sahip erkek hastaları iyileştirme oranının en az 0.80 olduğunu iddia etmektedir.  
Böyle bir durumda ortaya çıkacak I. Tip ve II. Tip hatayı ve testin gücünü tanımlayarak yorumlayınız. **(10 puan)**