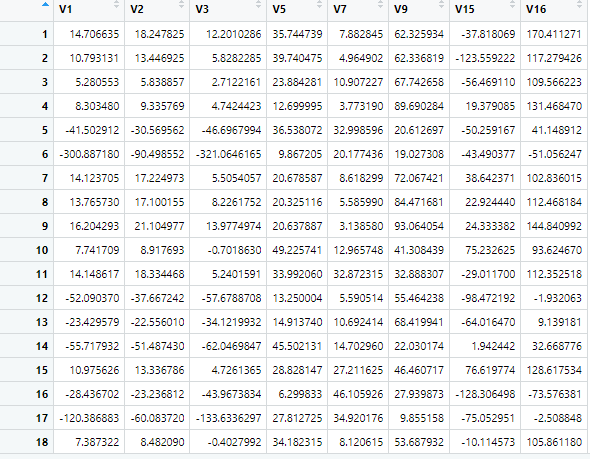
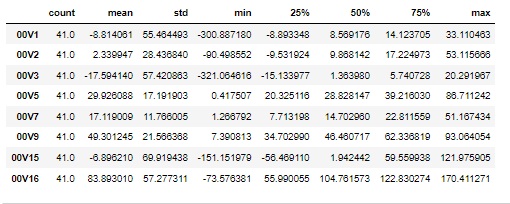
Hazırlayan: Ruveyda Nur ŞEN

Kullandığımız Programlar: Excel, R, Python, SPSS

41 bankaya ait değişkenler ve değerleri aşağıda verilmiştir. Bu veri üzerinden yaptığımız çeşitli analizler aşağıda yer almaktadır.

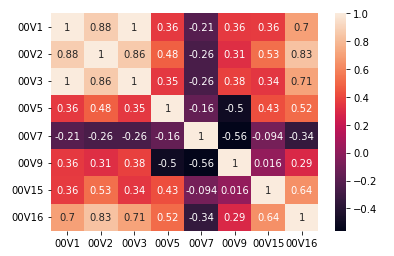


Tanımlayıcı istatistikler

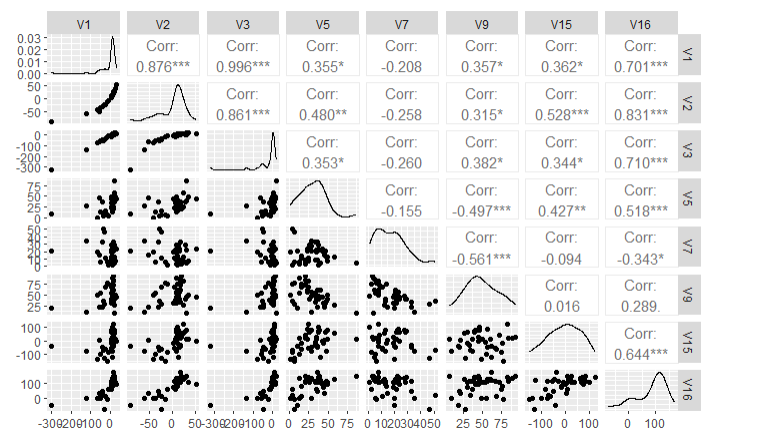


Tanımlayıcı istatistiklere baktığımızda ortalaması en büyük olan değişkenimiz V16 dır ve en düşük olan ise V3 değişkenidir. Mod, medyan ve ortalama birbirine ne kadar yakınsa o kadar normal dağıldığı varsayılır. Buna baktığımızda en normal dağılan V5 gibi görünüyor.

Aşağıda değişkenler arasındaki korelasyonları inceleyebileceğimiz bir sıcaklık haritası yer almaktadır. Korelasyon katsayısı -1 ile +1 arasında değerler alır. r= -1 ise tam negatif doğrusal bir ilişki vardır. r= +1 ise tam pozitif doğrusal bir ilişki vardır. r= 0 ise iki değişken arasında ilişki yoktur. +-0,50 ve +-0,69 arasındaki korelasyon değerleri orta güçlü; +-0,70 ve 0,89 arasındaki değerler yüksek güçlü; +-0,90 ile +-1 arasındaki değerler ise çok yüksek güçlü ilişkiyi ifade eder. Buradan V3 ve V1in korelasyonu 1’dir ve bu çok güçlü pozitif ilişki demektir. Ayrıca V2 ile V16 değişkenleri arasındaki 0,83 , V1 ile V2 arasındaki 0,88 , V3 ile V2 arasındaki 0,86 değerleri yüksek güçlü pozitif korelasyon ilişkisini gösterir. V7 ile V9 arasında dikkate alınabilecek -0,56 değerinde negatif orta düzey korelasyon vardır.



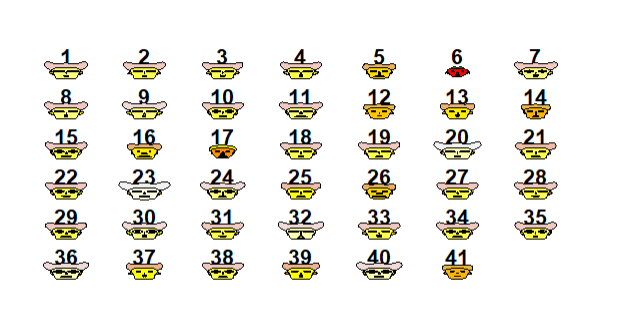
Matrix Şeklinde Scatter Plot



Saçılım grafiğimizden korelasyon değerlerine tekrar göz atabiliriz. Ayrıca yoğunluk grafiğini de inceleyerek normal dağılıp dağılmadığı hakkında yorum yapabiliriz. Buradan en normal dağılan V9 ve V15 değişkeni gibi gözükmektedir. V1, V2, V3 sola çarpık gibi gözükmektedir. Ayrıca normallikleri aşağıda test ettik.

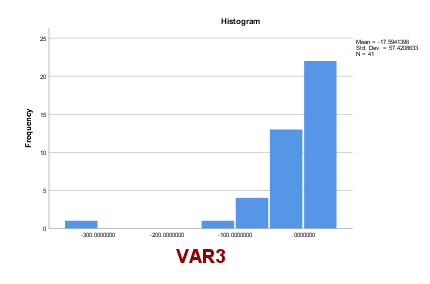
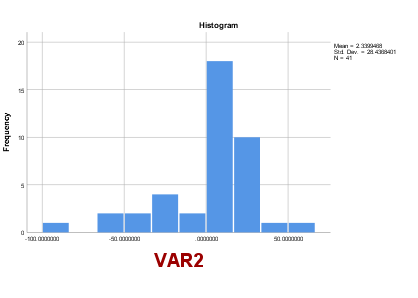
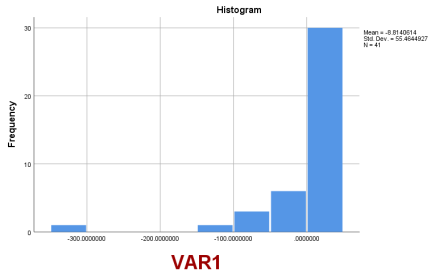
Chernoff yüzleri

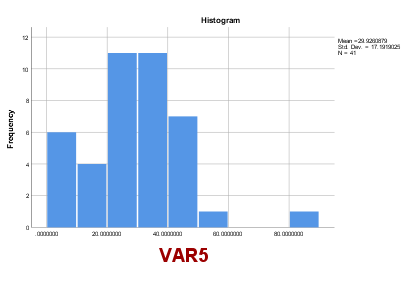
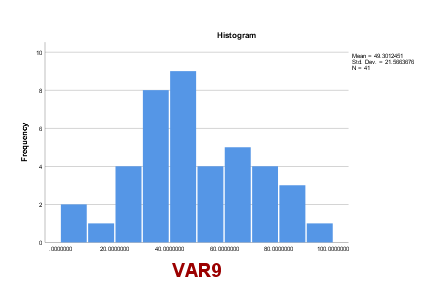
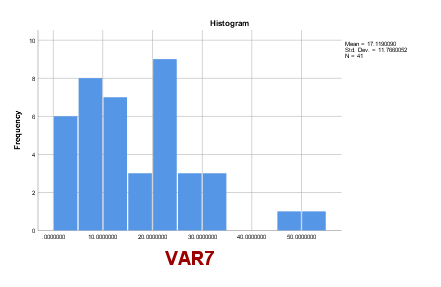
Chernoff Yüzleri Grafiği’ni incelediğimizde 6. yüz en farklı görünen yüzdür. Bu da 6. Bankanın değerlerinin normal dağılmadığını göstermektedir.

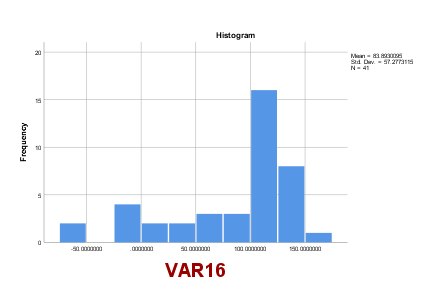
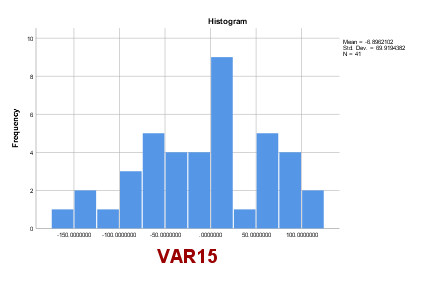


Histogram

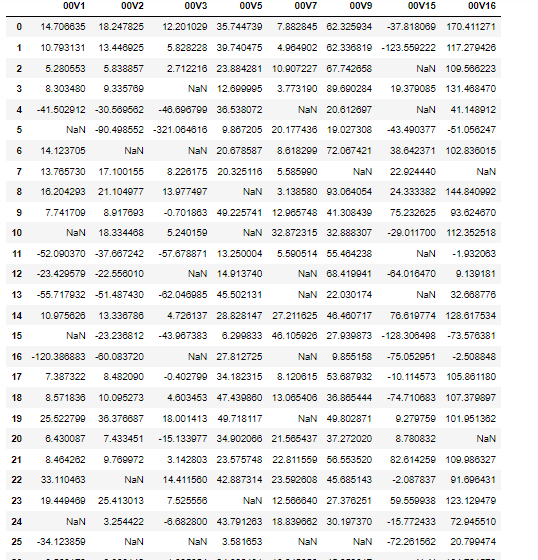
Değişkenler için ayrıca çizdirdiğimiz histogram grafiklerini incelediğimizde 1. ,3. ,16. Değişkenlere ait histogram grafiklerinin sola çarpık olduğu, 15. Ve 9. Değişkenlere ait histogram grafiklerinin normal dağıldığı yorumu yapılabilir.

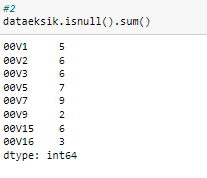


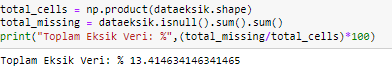
 



Verimizde eksik gözlem olmadığı için eksik gözlemli bir veri oluşturduk.





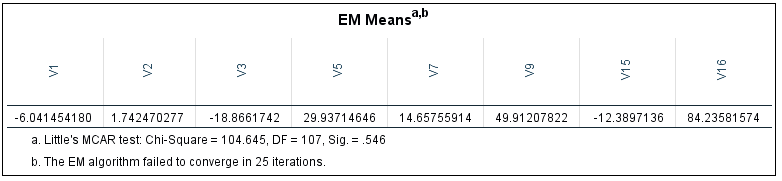


Yukarıda değişkenlerin sahip olduğu eksik veri sayıları gözükür. Ayrıca toplam eksik verinin, verinin %13.41’lik bir kısmını oluşturur.Eksik verilerin normal dağılıp dağılmadığıyla ilgili hipotezimiz ve analizimiz aşağıda verilmiştir.

H0: Eksik veriler rastgele dağılmıştır.

H1: Eksik veriler rastgele dağılmamıştır.

Little’s MCAR testi sonucuna göre Sig=0.546 değeri elde edilir. 0.546>0.05 olduğu için H0 hipotezi red edilemez. Eksik veriler rastgele dağıldığı sonucuna ulaşılır.

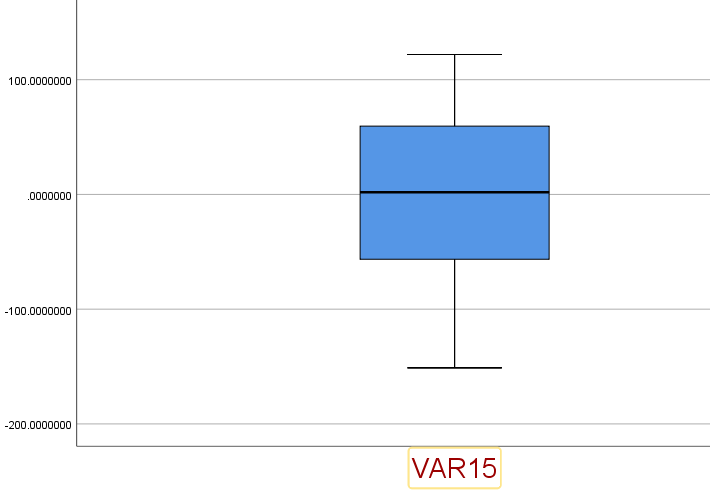
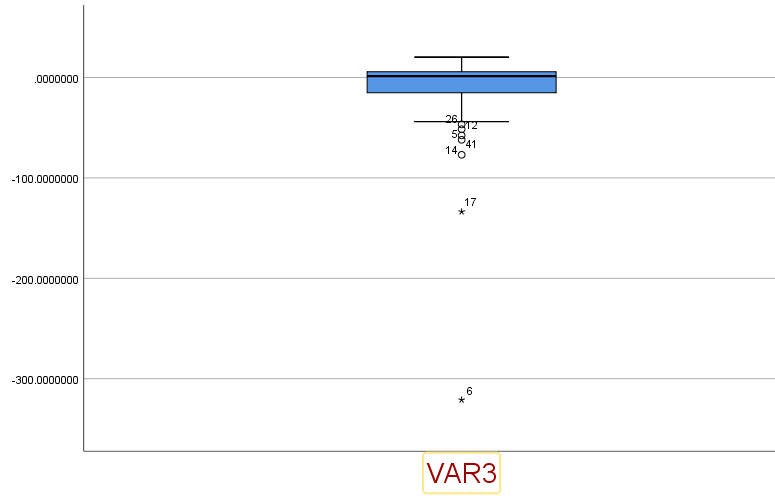


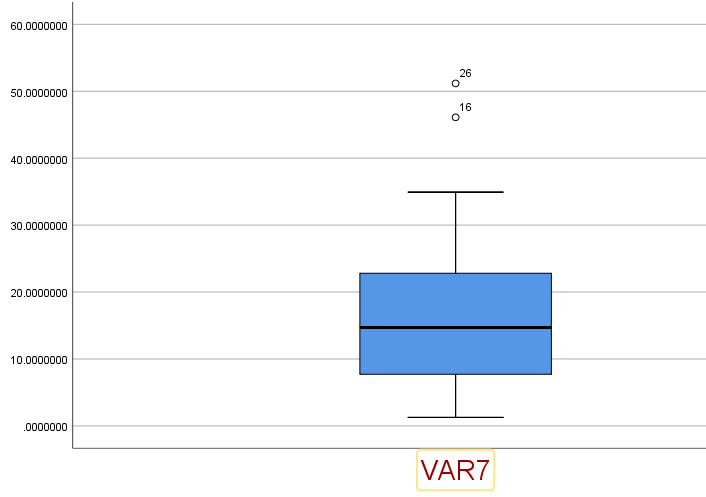
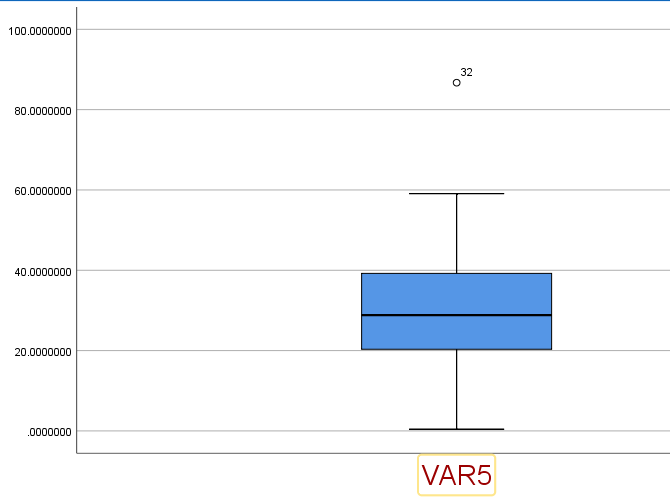
Eksik Veriler ve Çözüm yöntemleri

* Eksik veriden kaynaklanacak hata toleransını azaltmak için çok sayıda method vardır. En basit yöntem olarak eksik – kayıp veriyi yok sayarak diğer adıma geçmektir. Fakat eksik verinin çok olduğu yöntemlerde hata yüzdesi çok fazla olacağı için bu durum araştırmacının yükünü arttıracaktır. Bu yöntemlerin bazıları Hot / Cold Deck, EM (beklenti maksimizasyonu), regresyon, karar ağacı, Naive Bayes, K – en yakın komşuluk, ortalama değer ve en son gözlem gibi yöntemler kullanılabilir.

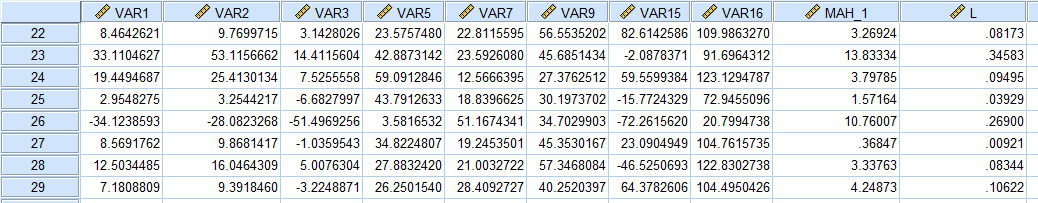
Box plotlar

Box-plotlar aşırı değerleri görmemizi sağlar. Aşağıda 4 değişkenin box-plot grafiği verilmiştir. Bunları seçerken hiç aşırı değer olmayan(V15), 7 tane aşırı değere sahip(V3), 1 tane aşırı değere sahip(V5) ve 2 tane aşırı değere sahip(V7) box plotları seçmeyi tercih ettik.





Aşırı değerleri çok değişkenli yaklaşımla kontrol ederken mahalonobis uzaklığını kullanabiliriz. L(ki kare) değeri 0.01’den küçükse gözlemin aykırı değer olduğunu söyleyebiliiz. Bizim verimizdeki 27. bankanın l değeri 0.00921 olduğu için aşırı değerdir.



Aykırı değerler için çözüm önerileri;

* Aykırı gözlemlerin silinmesi (Trimming)
* Aykırı gözlemlerin yerine değer ataması (Imputation)
* Aykırı gözlemlerin sınırlandırılması (Winsorization)
* Aykırı gözlemlerin dönüştürülmesi (Transformation)

Log dönüşümü

Karekök dönüşümü

* Aykırı gözlemlerin gruplanması (Binning)

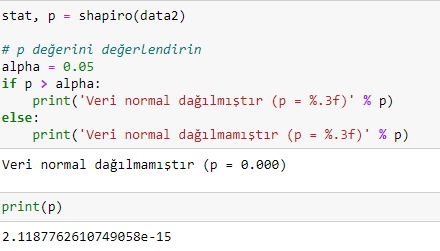
Verimizin normal dağılıp dağılmadığını kontrol etmek için Shapiro-Wilk testini kullandık. Hipotezlerimiz aşağıdadır.

H0: Verinin dağılımı normaldir.

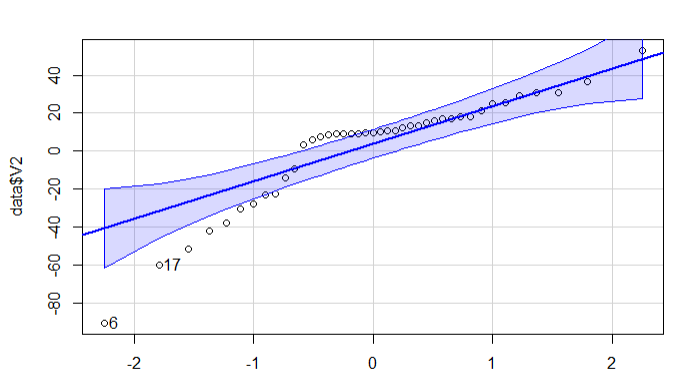
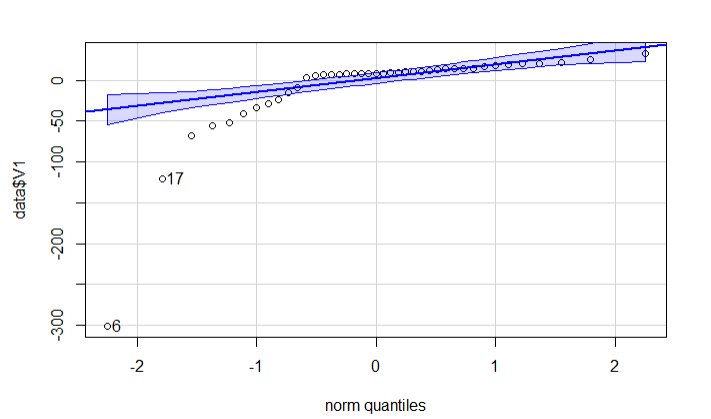
H1: Verinin dağılımı normal değildir.

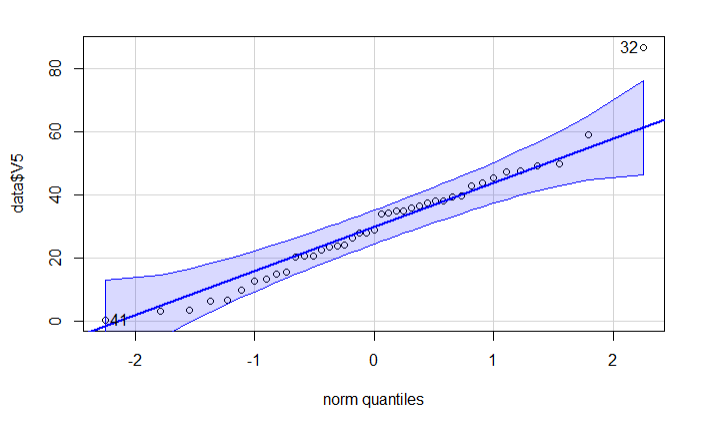
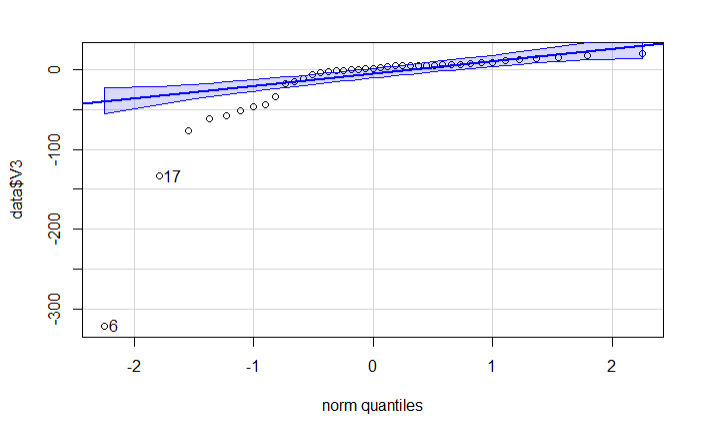
Uygulanılan test sonucunda p değeri 0,05’ten küçük olduğu için H0 hipotezi red edilir.

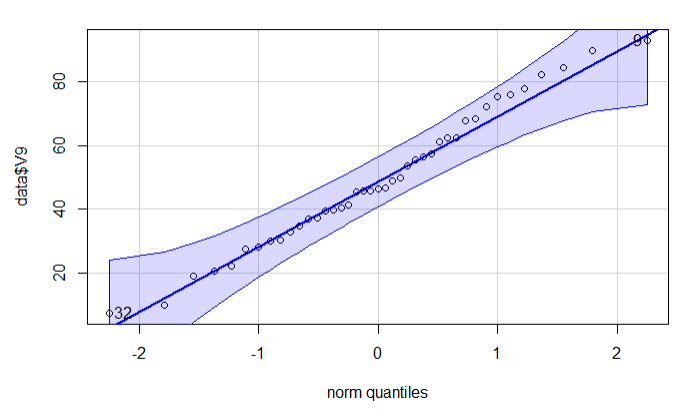
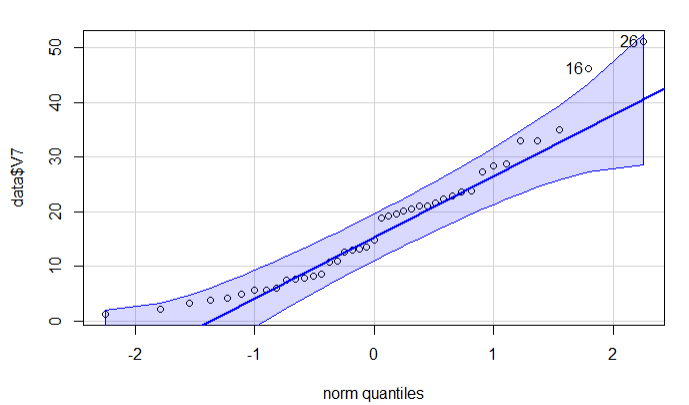
Verinin normal dağılmadığı sonucuna buradan ulaşılır.

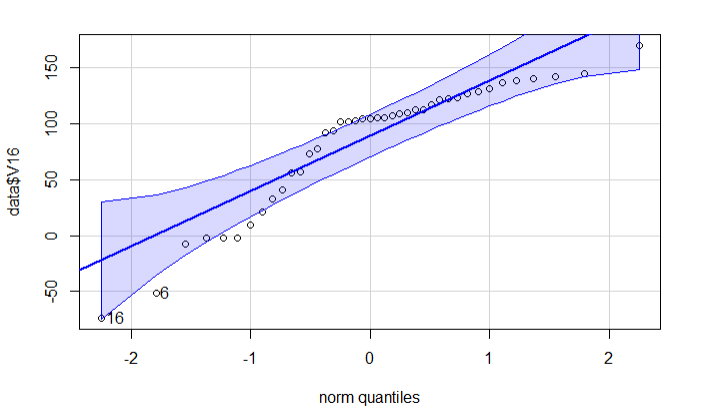
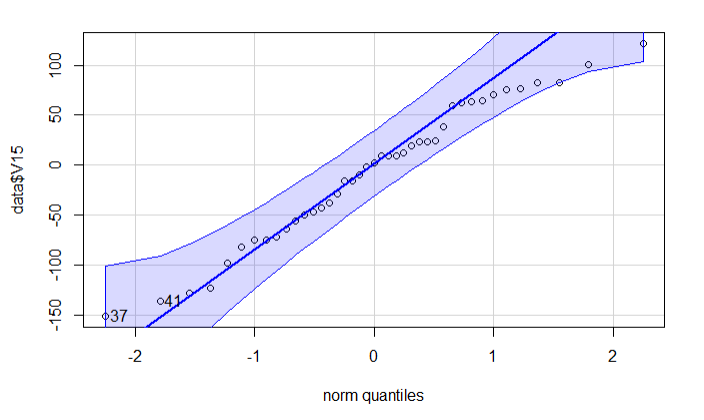


Verimizin içindeki değişkenlerin normal dağılıp dağılmadığına qq-plotlar yardımıyla önce inceleyip, sonrasında her birine Shapiro-Wilk testini uyguladık. Bu plotlardan V5, V15 ve V9 normal dağılıyor gibi gözüküyor.



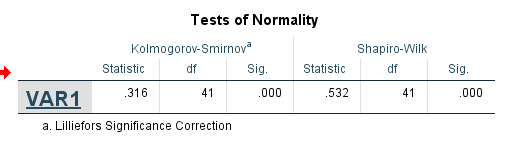
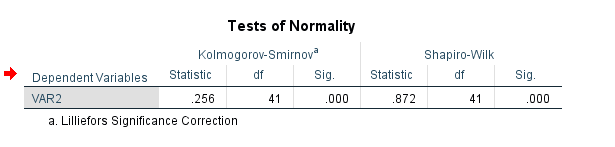


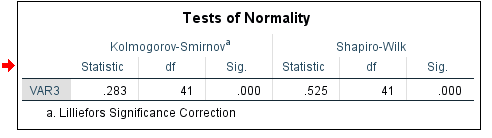


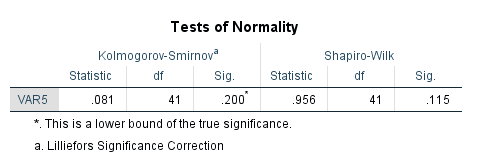


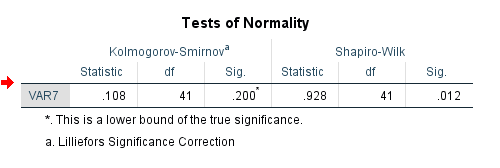
Aşağıda değişkenlerimizin normalliğini incelemek için yaptığımız Shapiro-Wilk test sonuçları yer almaktadır.

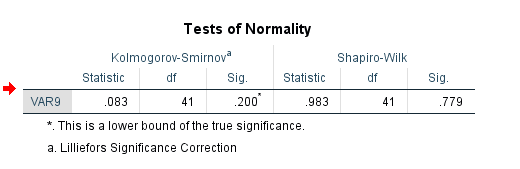
Sig değeri 0,05’ten büyük olan V5, V9, V15 değişkenleri normal dağılmıştır.

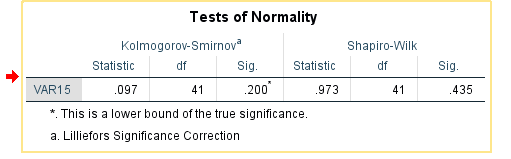
 

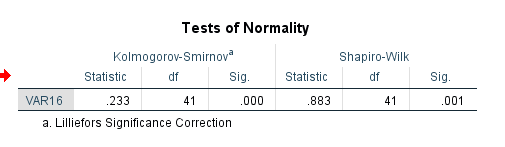






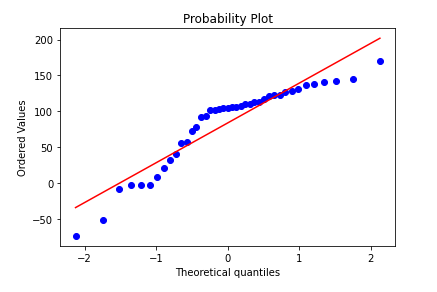


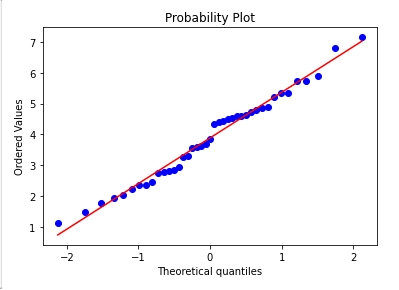




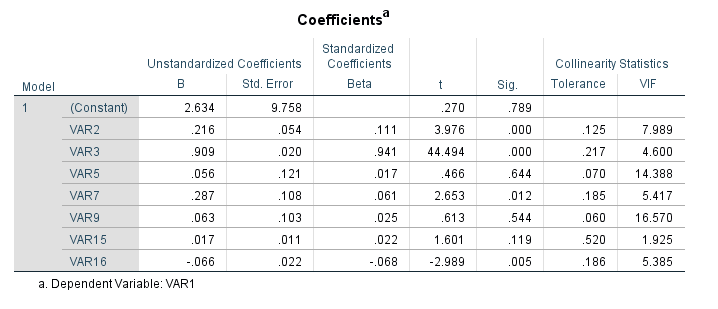
V7 için normallik dönüşümü yapmak için karekök dönüşümü yöntemini kullandık.

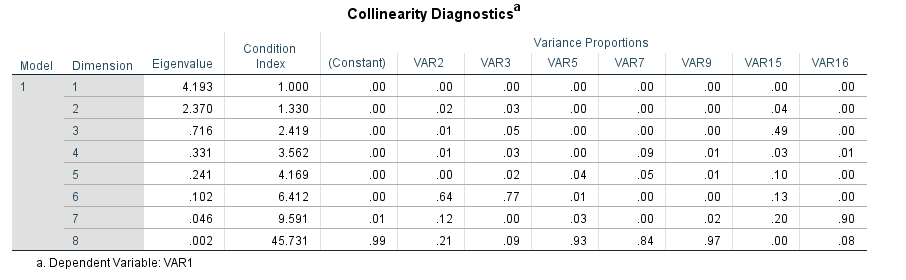
Normal dağılmayan değişkenlerden biri olan V7 değişkenine karekök dönüşümü uyguladık değişken 2. grafikte görüldüğü gibi normal dağılımlı hale gelmiştir.

 <- öncesi

<- sonrası

Aşağıda çoklu bağlantı problemini incelemek için SPSS’ten yaptığımız analiz çıktıları yer almaktadır.





Vif ve tolarance değerlerine bakarız. Tolarance> 0.20 ve VIF<10, Condition Index<30 olmalı.Çoklu bağlantı problemi yoksayılabilir.