**Bağımsız Değişkene Sahip Örneklem İçin Araştırma ve Analiz**

**İtalya’da 3 farklı şarap kültürü üzerine yapılan bir çalışma:**

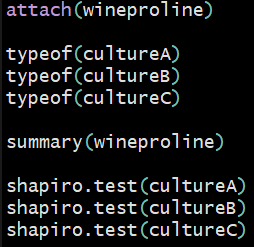
İtalya’da aynı bölgede fakat 3 farklı kültürde üretilen şaraplar için kimyasal araştırmalar yapılmıştır. Araştırmamızda şarapların içinde bulunan proline değerlerini karşılaştırarak şarapların yetiştiği kültür ve içinde bulunan proline maddesiyle ilişkisini araştırmak istiyoruz.

Araştırma için 3 farklı kültür ve 8 gözlem seçilmiştir. Şarapların içindeki proline gözlemleri aşağıdaki gibidir;

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.Culture(proline) | 1680 | 1547 | 1515 | 1510 | 680 | 735 | 760 | 770 |
| 2.Culture(proline) | 985 | 937 | 886 | 870 | 278 | 290 | 312 | 315 |
| 3.Culture(proline) | 880 | 855 | 840 | 835 | 415 | 425 | 470 | 480 |

**1)**

Bulduğumuz veri kümesine ait özetleyici istatistiklerin çıktıları için R programında kullanılan kodlar;

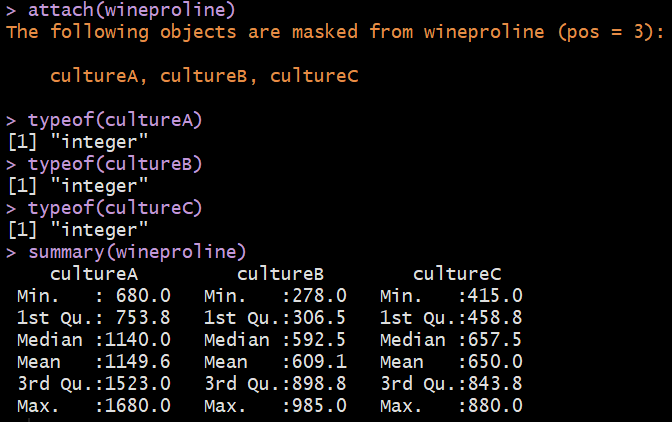


“Attach” komutunu kullanma nedenimiz veri setimiz içindeki değişkenleri R programına tanıtmak, 3 farklı kültürün üzerine gerekli kodlarla araştırmamızı yapabilmek içindir.

Kullandığımız kodlar bize verilerimiz hakkındaki özetleyici istatistikleri (summary), veri tiplerini (typeof) ve veri verilerimizin normal dağılmadığını, parametrik testlerin kullanılabileceğini gösteren (shapiro.test) kodlarıdır.

(shapiro wilks testi kullanmamızın amacı gözlem sayımızın <50 olmasıdır.)

Uygulanan kodlar sonucunda çıktılar;



Bütün değişkenlerimizin “integer” olduğu bilgisine eriştik.

**---** A kültürü için özetleyici istatistikleri, min. ve max. değerlerini (680 ve 1680 arasında), ortalamasını (1149.6), ortancasını (1140) ve çeyreklik değerlerini ( 1.çey=753.8 , 3.çey=1523)

kullanılan kodlar sonucunda elde ettik.

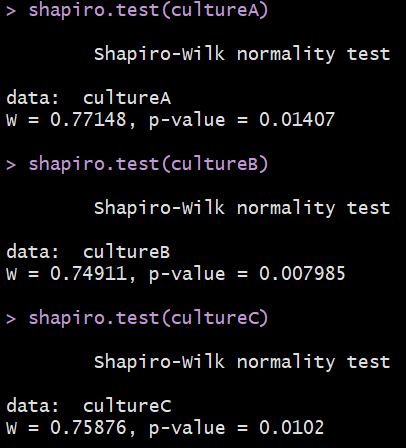
**---** B kültürü için özetleyici istatistikleri, min. ve max. değerlerini (278ve 985 arasında), ortalamasını (609.1), ortancasını (592.5) ve çeyreklik değerlerini ( 1.çey=306.5, 3.çey=898.8)

kullanılan kodlar sonucunda elde ettik.

**---** C kültürü için özetleyici istatistikleri, min. ve max. değerlerini (415ve 880 arasında), ortalamasını (650), ortancasını (657.5) ve çeyreklik değerlerini ( 1.çey=458.8 , 3.çey=843.8)

kullanılan kodlar sonucunda elde ettik.

3 kültüre de teker teker shapiro-wilks testi uygulandığında alınan çıktılar;



Elde edilen çıktıda üç kültür içinde gözlenen gözlemlerin normal dağılım sergilemediği

“p-değeri < α = 0.05” ifadesinden anlaşılmaktadır. Bu nedenle parametrik olmayan testler uygulanabilir.

**3)**

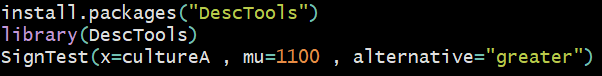
**Tek Örneklem İşaret Testi**

A kültürü içinde üretilen şaraplar için proline maddesi miktarının 1100’den büyük olduğunu test etmek için kurulan hipotezler;

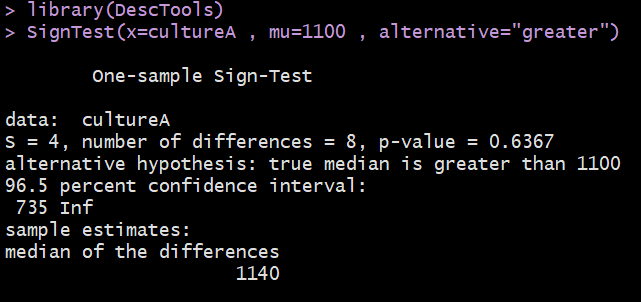
Ԩ0 : Ɵ = 1100

Ԩ1: Ɵ > 1100

Bu hipotezleri test etmek için kullanılan R kodları;



Kodlar kullanıldıktan sonra işaret testi çıktısı;



p-değeri = 0.6367 > α = 0.05 Ԩ0  Kabul.

**Yorum =** Bu çıktıya göre A kültürü içerisinde üretilen şaraplardaki proline maddesi miktarının ortanca bakımından 1100’e eşit olduğunu %5 anlamlılık düzeyinde söyleyebiliriz.

Bir fikir edinebilmek adına bu işaret testi için bir hipotez daha kuruyorum;

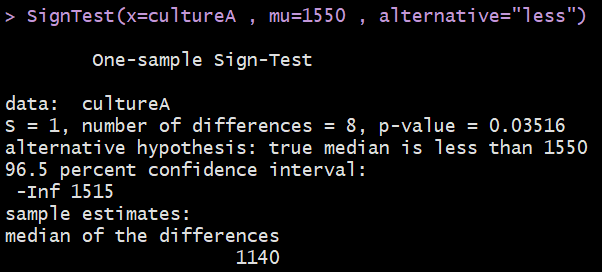
Ԩ0 : Ɵ = 1550

Ԩ1 : Ɵ < 1550

Bu hipotezleri test etmek için kullanılan R kodu;



Kod kullanıldıktan sonra işaret testi çıktısı;



p-değeri = 0.03516 < α = 0.05 Ԩ0  Ret.

**Yorum =** Bu çıktıya göre A kültürü içerisinde üretilen şaraplardaki proline maddesi miktarının ortanca bakımından 1550’den küçük olduğunu %5 anlamlılık düzeyinde söyleyebiliriz.

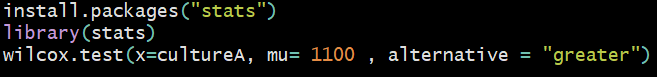
**Tek Örneklem Wilcoxon İşaret Sıra Sayıları Testi**

A kültürü içinde üretilen şaraplar için proline maddesi miktarının 1100’den büyük olduğunu test etmek için kurulan hipotezler;

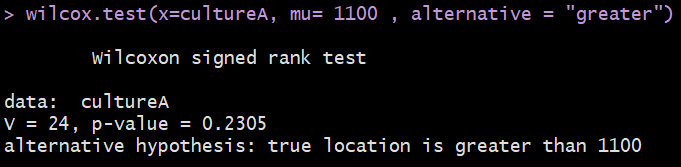
Ԩ0 : Ɵ = 1100

Ԩ1: Ɵ > 1100

Bu hipotezleri test etmek için kullanılan R kodları;



Kodlar kullanıldıktan sonra işaret testi çıktısı;



p-değeri = 0.2305 > α = 0.05 Ԩ0  Kabul.

**Yorum =** Bu çıktıya göre A kültürü içerisinde üretilen şaraplardaki proline maddesi miktarının ortanca bakımından 1100’e eşit olduğunu %5 anlamlılık düzeyinde söyleyebiliriz.

Bir fikir edinebilmek adına bir hipotez daha oluşturuyorum.

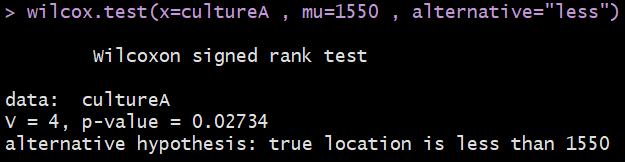
Ԩ0 : Ɵ = 1550

Ԩ1 : Ɵ < 1550

Bu hipotezleri test etmek için kullanılan R kodu;



Kodlar kullanıldıktan sonra işaret testi çıktısı;



p-değeri = 0.02734 < α = 0.05 Ԩ0  Ret.

**Yorum =** Bu çıktıya göre A kültürü içerisinde üretilen şaraplardaki proline maddesi miktarının ortanca bakımından 1550’den küçük olduğunu %5 anlamlılık düzeyinde söyleyebiliriz.

**4)**

**Mann-Whitney Testi**

Man-Whitney U testi bağımsız iki örneğin aynı ana kütleden alınmış rasgele örnekler olup olmadığını test etmek için kullanılan testlerdir.

Gerekli hipotezler;

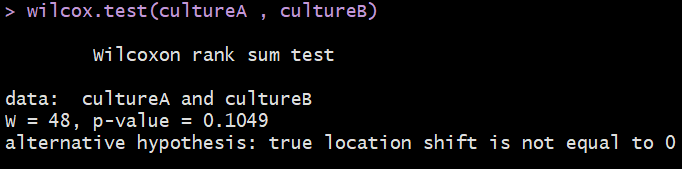
Ԩ0 : Ma = Mb  (İki örnek aynı ana kütleden gelmektedir. Ortalamalar eşittir.)

Ԩ1 : Ma ≠ Mb  (İki örnek farklı ana kütleden gelmektedir. Ortalamalar eşit değildir.)

Kurulan hipotezleri test etmek için kullanılan R kodu;



Kod kullanıldıktan sonra alınan R çıktısı;



p-değeri = 0.1049 > α = 0.05 Ԩ0  Kabul.

Alınan çıktıya göre A kültürü ve B kültürünün aynı ana kütleden ( konum yönünden) geldiğini %5 anlamlılık düzeyinde söyleyebiliriz.

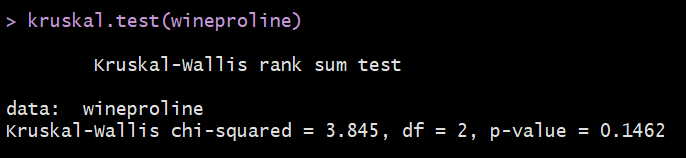
**6)**

**Kruskal-Wallis Testi**

Gerekli hipotezler;

Ԩ0 : Şarapların içindeki proline maddesi miktarı bakımından kültürler arasında konum yönünden fark yoktur.

Ԩ1 Şarapların içindeki proline maddesi miktarı bakımından kültürler arasında konum yönünden fark vardır.



p-değeri = 0.1462 > α = 0.05 Ԩ0  Kabul.

Alınan çıktıya göre şarapların içindeki proline maddesi miktarı bakımından kültürler arasında konum yönünden fark olmadığını %5 anlamlılık düzeyinde söyleyebiliriz.

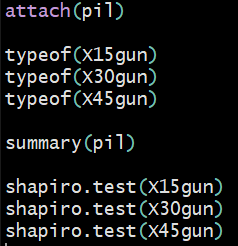
**Bağımlı Değişkene Sahip Örneklem İçin Araştırma ve Analiz**

A marka kalem pilin rafta bekleme süresine ilişkin bir araştırma yapılmak isteniyor. Bunun için rafta bekleme süresi değişkenleri için 15 gün, 30 gün ve 45 gün olmak üzere 3 değişken belirlenmiştir. Bunlar için 6 gözlem alınmıştır.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **15Gün** | 6.5 | 7.6 | 41.1 | 185.9 | 30.6 | 29.9 |
| **30Gün** | 90.3 | 30.1 | 39.4 | 37.6 | 26.5 | 17.1 |
| **45Gün** | 24.1 | 29.9 | 32.1 | 22.4 | 90.4 | 51.2 |

**2)**

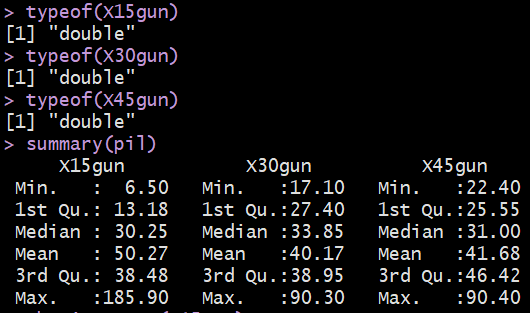
Bulduğumuz veri kümesine ait özetleyici istatistiklerin çıktıları için R programında kullanılan kodlar;

****

Kullandığımız kodlar bize verilerimiz hakkındaki özetleyici istatistikleri (summary), veri tiplerini (typeof) ve veri verilerimizin normal dağılmadığını, parametrik testlerin kullanılabileceğini gösteren (shapiro.test) kodlarıdır.

(shapiro wilks testi kullanmamızın amacı gözlem sayımızın <50 olmasıdır.)

Uygulanan kodlar sonucunda çıktılar;



Bütün değişkenlerimizin “double” olduğu bilgisine eriştik.

**---** 15 gün rafta beklemiş pil için özetleyici istatistikleri, min. ve max. değerlerini (6.5 ve 185.90 arasında), ortalamasını (50.27), ortancasını (30.25) ve çeyreklik değerlerini ( 1.çey=13.18 , 3.çey=38.48)

kullanılan kodlar sonucunda elde ettik.

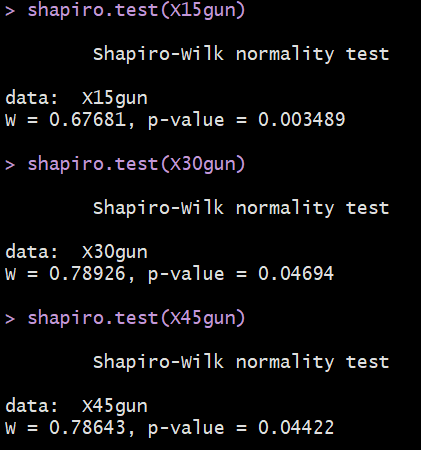
**---** 30 gün rafta beklemiş pil için özetleyici istatistikleri, min. ve max. değerlerini (17.10 ve 90.3 arasında), ortalamasını (40.17), ortancasını (33.85) ve çeyreklik değerlerini ( 1.çey=27.4, 3.çey=38.95)

kullanılan kodlar sonucunda elde ettik.

**---** 45 gün rafta beklemiş pil için özetleyici istatistikleri, min. ve max. değerlerini (22.4 ve 90.4 arasında), ortalamasını (41.68), ortancasını (31) ve çeyreklik değerlerini ( 1.çey=25.55 , 3.çey=46.42)

kullanılan kodlar sonucunda elde ettik.

3 değişkene de teker teker shapiro-wilks testi uygulandığında alınan çıktılar;



Bütün değişkenlerin 0.05 değerinden küçük olduğunu bu nedenle normal dağılmadığını söyleyebiliyoruz. Böylelikle parametrik olmayan yöntemler kullanılabileceğini görmüş olduk.

**5)**

**Wilcoxon İşaret Sıra Sayıları Testi**

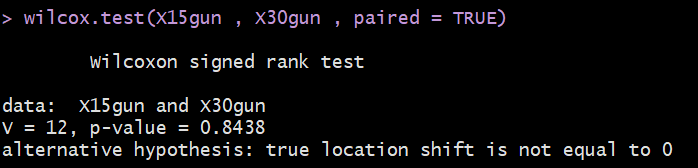
Bağımlı iki grup için;

Gerekli hipotezler;

Ԩ0 : Q15 = Q30  (15 Gün ve 30 Gün rafta bekleme sürelerine ilişkin piller arasında konum yönünden fark yoktur.)

Ԩ1 : Q15 ≠ Q30  (15 Gün ve 30 Gün rafta bekleme sürelerine ilişkin piller arasında konum yönünden fark vardır.)

Gerekli kodlar uygulandıktan sonra alınan çıktı;



p-değeri = 0.8438 > α = 0.05 Ԩ0  Kabul.

15 Gün ve 30 Gün rafta bekleme sürelerine ilişkin piller arasında konum yönünden fark olmadığını %5 anlamlılık düzeyinde söyleyebiliriz.

**7)**

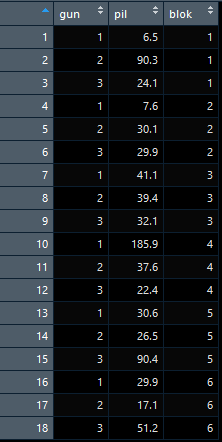
**Friedman Testi**

Gerekli hipotezler;

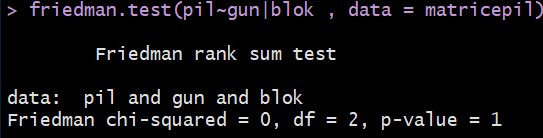
Ԩ0 : Pillerin rafta bekleme süreleri için yayılım yönünden gruplar arasında fark yoktur.

Ԩ1 : Pillerin rafta bekleme süreleri için yayılım yönünden gruplar arasında fark vardır.

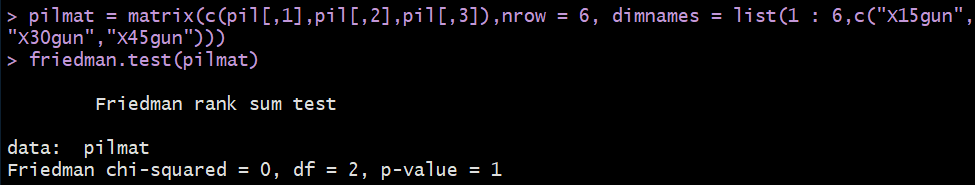
Blokları koymak için elimizdeki veri setine blok ekledik.



Aşağıdaki kodu kullanarak çıktıya eriştik;



p-değerinin 1 gelmesi nedeniyle bir de matris yaratarak denemek istedim.



Bunun sonucunda da aynı değeri aldım ve emin oldum. Ԩ0 hipotezi kabul edildi.

Pillerin rafta bekleme süreleri için yayılım yönünden gruplar arasında fark olmadığını %5 anlamlılık düzeyinde söyleyebiliriz.

Kaynakça: İstatistiksel Yöntemlere Giriş Dr. Haydar Demirhan Dr. Canan Hamurkaroğlu Hacettepe Üniversitesi Yayınları 2011

Kaynakça: <https://archive.ics.uci.edu/ml/datasets/Wine>