

Ödev 1

(25 Ocak'ta düzenlendi)

A. Yetkin Alıcı, Elif Bozlak, Evrim Fer, Erinç Yurtman

22 Ocak 2018

Kısa olan kursun verimli geçebilmesi adına, lütfen 2 Şubat saat 17.00'ye kadar burada yazanları yapın ve yazdığımız kodları, AdımızSoyadımız.R şeklinde bir dosyaya kaydederek popgensim@gmail.com adresine gönderin. Notlarda ve ödevde anlaşılmayan yerler varsa aynı e-mail adresine sorularınızı iletmekten çekinmeyin.

R ve RStudio programlarını ders notunda yazan adresten indirip, bilgisayarınıza kurun. Sonrasında 1. ders notlarını, örnekleri kendiniz de RStudio ile tekrarlayarak okuyun. Özellikle mantıksal operatörler kısmının ve vektörden/matristen alt vektör/matris seçme kısmının üstünde durarak örnekleri anladığınızdan emin olun.

`length()`, `dim()`, `sort()`, `set.seed()`, `sum()`, `mean()`, `sd()`, `summary()`, `table()`, `plot()`, `hist()`, `boxplot()`, `rnorm()`, `runif()`, `substr()`, `gsub()`, `paste()` fonksiyonlarını, yardım dosyalarına bakarak ve kendiniz deneyerek nasıl çalıştıklarına aşina olun.

Aşağıdaki örneği okuduktan sonra, 3 problem için de yazdığımız kodları aynı dosyaya (AdımızSoyadımız.R) kaydederek bize gönderin. Sadece kodları göndermeniz yeterli, sonuçları ya da grafikleri göndermenize gerek yok.

Örnek Problem:

Tek lokusta, iki alelli bir özelliğin, her alelin birer gamet ürettiği varsayımı altında diploid aktarım simülasyonu: 100 bireyden oluşan ve bireylerin ("AA", "Aa", "aa") genotiplerine sahip olma olasılıklarının eşit olduğu bir başlangıç populasyonunda, alel ve genotip frekanslarında bir nesilde olan değişimi, aşağıdaki adımları izleyerek gösterin.

- Eleman olarak "AA", "Aa" ve "aa" karakterlerinden oluşan, "Genotipler" adında bir vektör oluşturun.

```
Genotipler = c("AA", "Aa", "aa")
```

- "Populasyon1" adında, 3 satır ve 100 sütundan oluşan, boş bir karakter matrisi oluşturun. Data olarak boş karakter girerek bu yapılabilir.

```
Populasyon1 = matrix(data = "", nrow = 3, ncol = 100)
```

- Genotipler vektöründen, rastgele 100 eleman seçerek, Populasyon1 matrisinin ilk satırına bu elemanları yerleştirin.

```
Populasyon1[1,] = sample(Genotipler, 100, replace = T)
```

- Populasyon1 matrisinin 2. ve 3. satırlarına aynı sütundaki 1. satırda yazan genotipin içerdiği alelleri ayrı ayrı yazalım. Bunun için, `substr()` fonksiyonunu kullanabiliriz. Bu fonksiyon, ilk parametre olarak verdiğimiz karakter değerinin; 2. ve 3. parametre olarak verdiğimiz sayılar arasında bulunan parçasını geri döndürür.

- İlk parametre olarak Populasyon1 matrisinin ilk satırını verirse, işlemi her elemana ayrı ayrı uygulayıp aynı boyda bir satır geri verecektir.
- Populasyon1 matrisinin 2.satırına, genotiplerin ilk alellerini yazdırmak için, substr() fonksiyonunu kullanın. Aynı işlemi 2.alel için tekrarlayarak sonucu Populasyon1 matrisinin 3. satırına atayın.

```
Populasyon1[2,] = substr(Populasyon1[1,], 1,1)
Populasyon1[3,] = substr(Populasyon1[1,], 2,2)
```

Populasyon1 matrisinin nasıl görüldüğüne bakmak için, küçük bir alt matrisini yazdıralım:

```
Populasyon1[,1:8]
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,] "AA" "aa" "aa" "AA" "AA" "aa" "Aa" "aa"
## [2,] "A"  "a"  "a"  "A"  "A"  "a"  "A"  "a"
## [3,] "A"  "a"  "a"  "A"  "A"  "a"  "a"  "a"
```

- “Populasyon2” adında, 3 satır ve 100 sütundan oluşan, boş bir karakter matrisi daha oluşturun.

```
Populasyon2 = matrix(data = "", nrow = 3, ncol = 100)
```

- Populasyon2 matrisinin 2. ve 3. satırına Populasyon1 matrisinin aynı satırlarındaki alelleri rastgele yerleştirin. Her alelin tek gamet oluşturduğu varsayımından dolayı, seçilen alelin tekrar seçilemeyeceği şekilde (without replacement) seçim yapılacak, her alel yeni populasyonda da tam bir kez yer bulacak.

```
Populasyon2[2,] = sample(Populasyon1[2,], 100, replace = FALSE)
Populasyon2[3,] = sample(Populasyon1[3,], 100, replace = FALSE)
```

- Populasyon2 matrisinin 1. satırına, 2. ve 3. satırlardaki alelleri birleştirerek genotipleri yazdırın. paste() fonksiyonu, virgülle ayrılmış parametre olarak verilen karakter değerlerini, her birinin arasına sep = “” parametresinde tırnak içindeki karakter gelecek şekilde birleştirip tek eleman halinde geri verir. Tırnak işaretinin içi boşsa karakterler bitişik olacak şekilde birleştirilir.

```
Populasyon2[1,] = paste(Populasyon2[2,],Populasyon2[3,], sep = "")
```

Populasyon2 matrisinin nasıl görüldüğüne bakmak için, küçük bir alt matrisini yazdıralım:

```
Populasyon2[,1:8]
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,] "aa" "Aa" "aa" "aa" "aa" "Aa" "Aa" "AA"
## [2,] "a"  "A"  "a"  "a"  "a"  "A"  "A"  "A"
## [3,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "A"
```

- Populasyon1 ve 2’deki allel ve genotip sıklıklarını, table() fonksiyonunu kullanarak gösterin ve birer değişken altında kaydedin. Bu fonksiyon, parametre olarak verilen verideki her bir değerin veride kaç kez geçtiğini tablo halinde sunar.

```
Alel_dagilimi1 = table(Populasyon1[2:3,])
Genotip_dagilimi1 = table(Populasyon1[1,])
Alel_dagilimi2 = table(Populasyon2[2:3,])
Genotip_dagilimi2 = table(Populasyon2[1,])
```

- Kodlarınızı bir dosyada arka arkaya yazarken, arada açıklamalar yazmak kodun başkaları tarafından daha iyi anlaşılabilmesi için gerekli oluyor. “#” sembolü ile başlayan satırları R dikkate almıyor. Her birkaç satır öbeğinin üstüne kısa bir açıklama yazmayı (en azından problemleri “# Problem X” satırlarıyla ayırmak gibi) deneyebilirsiniz.

Problem 1:

Tek lokusta, iki alelli bir özelliğin, her alelin sonsuz gamet ürettiği varsayımı altında diploid aktarım simülasyonu: 100 bireyden oluşan ve bireylerin (“AA”, “Aa”, “aa”) genotiplerine sahip olma olasılıklarının eşit olduğu başlangıç populasyonunda, alel ve genotip frekanslarında bir nesilde olan değişimi gösterin.

- Örnek problemdeki adımları, Populasyon2’nin alellerini oluşturma adımına kadar uygulayın.
- Sonsuz gamet varsayımından dolayı, Populasyon2’nin alellerini oluşturma adımı farklı olacak.
- Bu sefer, Populasyon1’in 2. ve 3. satırlarındaki alellerin toplamından oluşacak tek bir gen havuzu oluşturun. Bu havuz vektör olmak zorunda değil, bir matris de olabilir ve bu matristen yeni populasyonun alellerini seçebiliriz. Populasyon1’in 2. ve 3. satırlarından ve tüm sütunlarından oluşan bir alt matris seçin ve bunu “Gametler” adında bir değişkende tutun.
- sample() fonksiyonunu kullanarak, Gametler matrisinden, seçilen alelin tekrar seçilebileceği şekilde (with replacement) 100 alel seçerek Populasyon2’nin 2. satırına yazın. Sonra aynımsı Populasyon2’nin 3.satırını için de yapın.
- Örnek problemde olduğu gibi, genotip satırını (Populasyon2 matrisinin 1.satırını) da oluşturun.
- Alel ve genotip frekanslarını hesaplatın.

Not 1: sample() fonksiyonu tek satırlık bir matris parçası ya da vektör almak zorunda değil, birden çok satırdan oluşan matris parçaları üstünde de çalışabilir.

Not 2: seçilen alelin tekrar seçilebileceği şekilde seçim yapmak, sonsuz gamet varsayımını simüle etmiş oluyor.

Problem 2:

Tek lokusta, iki alelli bir özelliğin, her alelin sonsuz gamet ürettiği ve “Aa” ve “aa” genotiplerinin %50 dezavantajlı olduğu durumda diploid aktarım simülasyonu: 200 bireyden oluşan ve bireylerin %50 olasılıkla heterozigot, %25’er olasılıkla da “AA” ya da “aa” genotiplerine sahip oldukları bir başlangıç populasyonunda, alel ve genotip frekanslarında bir nesilde olan değişimi, aşağıdaki adımları izleyerek gösterin.

- Örnek probleme göre farklılık, başlangıç populasyonunu oluşturmada ve 2. jenerasyon populasyonun alellerini belirlemede olacak.
- Örnek problemdeki adımları, parametreler yukarıdaki koşullara uyacak şekilde değiştirerek, Populasyon2’nin alellerini oluşturma adımına kadar uygulayın.
- 2.populasyonu oluşturacak olan gamet havuzunda, “AA” bireylerin diğer bireylerin 2 katı kadar temsil edilmesini istiyoruz.
- Öncelikle, 1.populasyonun matrisinin ilk satırını karşılaştırma operatörü kullanarak “AA” değeriyle karşılaştırm. Çıkan vektörü, istediğiniz isimde bir değişkende tutun, örneğin: “fit_olanlar”.

- Sadece fit olanların alellerinden oluşan gametleri bir “G1” matrisinde tutun. Bunun için, fit_olanlar vektörünü kullanarak, Populasyon1 matrisinin 2. ve 3. satırlarından ve fit olan sütunlarından oluşan bir alt matris seçin ve bu seçimi G1 değişkeninde tutun.
- “AA” olan bireyler dahil bütün bireylerin alellerinden oluşan bir matrisi de bir “G2” matrisinde tutun. Yani Populasyon1 matrisinin 2. ve 3. satırları ve tüm sütunlarından oluşacak bir alt matris seçin ve bu seçimi de G2 değişkeninde tutun.
- Sonrasında, G1 ve G2 matrislerini, cbind() fonksiyonuyla birleştirerek “Gametler” isimli bir matris oluşturun.
- Populasyon2 matrisini, Problem1’deki gibi oluşturun.
- Alel ve genotip frekanslarını hesaplatın.

Not: “AA” genotipleri gamet havuzuna 2 kere alel gönderirken, diğer genotipler gamet havuzuna 1 kez alel göndererek sonraki nesilde temsil açısından %50 dezavantajlı oluyorlar.

Problem 3:

Fenotip dağılımı simülasyonu: Nicel bir fenotipin dağılımının bir nesildeki değişimini, aşağıdaki adımları izleyerek simüle edin:

- Ortalaması 50, standart sapması 5 olan ve rassal (rastgele, random) normal dağılım gösteren bir fenotip için; 100 bireylik bir başlangıç populasyonu oluşturun. rnorm() fonksiyonu bu şekilde rassal değerler üretir. Bu fenotip değerleri bir vektör değişkeninde tutulsun.
- summary() fonksiyonu ile fenotip dağılımına dair istatistikleri gösterin ve hist() fonksiyonu ile histogramını çizin.
- Bireyleri (vektördeki değerleri) sort() fonksiyonuyla sıralayın ve en küçük fenotipe sahip yarısını seçerek oluşturacağınız başka bir vektör değişkeninde tutun.
- Bu seçtiğiniz yarısından, şu koşula uyan bir yeni populasyon üretmek istiyoruz: her birey yeni popülasyona iki birey katacak ve yeni bireylerin fenotipik değeri bir istatistiksel dağılımdan gelecek. Bu istatistiksel dağılım da, ortalaması atasal bireyin taşıdığı fenotip değeri ve standart sapması 5 olan normal dağılım olacak.
- Bu aşamada, fonksiyonların sadece tek bir değer üzerinde değil, bir vektörün bütün değerleri üzerinde işlem yapmasının bir örneğini göreceğiz. rnorm() fonksiyonuyla rassal değer üretirken, ortalama ya da standart sapma değeri olarak bir sayı değil de bir vektör verecek olursak, verdiğimiz vektördeki her değeri sırayla kullanıyor. Örneğin 50 sayı üretmek istiyorsak (ilk parametre 50 olacaksa), ve ortalama parametresi (2. parametre) olarak 50 değerden oluşan bir vektör verdiğimizde ve standart sapma olarak da tek bir değer verdiğimizde, vektördeki her bir değer bir kez ortalama değer olarak kullanılacak şekilde 50 rassal değer üretilir. Eğer üretilecek rassal değer sayısı, ortalama olarak kullanılacak vektörün uzunluğundan fazlaysa, vektördeki sayılar birer kez kullanıldıktan sonra başa dönülür ve tekrar kullanılırlar. Örneğimizde, 100 birey oluşturmak için (100 rassal fenotip değeri oluşturmak için) 50 bireyi (50 ortalama fenotip değerini) 2şer kez kullanacağız.
- rnorm() fonksiyonunu az önce anlatılan özelliğiyle kullanarak, ilk populasyonun en küçük fenotipli 50 bireyinden, 100 bireylik yeni bir populasyon üretilsin.
- Yeni popülasyon için de, summary() fonksiyonu ile fenotip dağılımına dair istatistikleri gösterin ve hist() fonksiyonu ile histogramını çizin.