**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Hesaplamalı Biyolojiye Giriş Dersi- 1.Ödev Raporu**

**Batuhan Şimşek --- 20360859055**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Öncelikle amacım, Bacteriophage Lambda isimli canlının 48502 BP uzunluklu genomunu analiz etmek,Poisson dağılımını kullanarak matematiksel bir model ile bir beklenti hesabı yapmak,sonrasında ise yaptığım gözlem sonuçları ile hesapladığım beklentiyi Ki-Kare(Chi-Square) hesabı ile karşılaştırmaktı.Program,MATLAB’a ait fonksiyonlar hariç,ders kitabımızdaki notlarda paylaşılan iki,kendi yazdığım bir fonksiyonu kullanarak çalışmaktadır.

***getGenBank() Fonksiyonu:***

MATLAB’e ait Bioinformatics Toolbox içerisinde bulunan “getgenbank()” fonksiyonu ile GenBank sitesinden Bacteriophage Lambda isimli canlının gen sekansını çektim.Aradığım RS bölgesini tanımladım ve strfind() fonksiyonu ile sekans içerisinde RS bölgesinin başlangıç noktalarını arattım. Daha sonra bulduğum sonuçları “fragmanlar” isimli bir yapıya aktardım. Fragman uzunluklarını ise ayrı bir “uzunluk” dizisine atadım.Daha sonra ise bulduğum uzunluk aralıklarını bir histogram ile ekranda gösterdim.Tüm bu işlemleri tek bir getGenBank() fonksiyonunda topladım.

***integralExpectedValue(i) Fonksiyonu:***

Teorik olasılıkları Poisson dağılımına göre hesaplamak için, “integralExpectedValue(i)” isimli fonksiyonu kullandım. P olasılık değeri olarak kaynak kitabımızın 88. Sayfasındaki “Table 3.1” den yararlanarak “0.00386” değerini esas aldım. Ayrıca fonksiyonun son aşamasında, hesapladığım olasılıkları, getGenBank () fonksiyonunda hesapladığım toplam fragman sayısı olan 328 ile çarparak istenilen aralık için teorik fragman sayısını hesaplamış oldum.Son aralığı fonksiyona uyarlamak sorun olacağı için onu ayrıca hesapladım.

getGenBank() fonksiyonunda çizdirdiğim histogramdan yararlanarak her aralık için gözlemlenmiş fragman sayılarını hesaplayarak ObsVal[] isimli bir diziye atadım,kodun devamında gözlemlenen değerleri bu dizi üzerinden çağırdım.Sonrasında her bir aralık için teorik ve gözlemlenen değerleri yazdırmaya yarayan kodları yazdım.

***chiSquare(Observed,Expected) Fonksiyonu:***

Bu fonksiyon kendisine gözlem ve beklenti sonuçlarını parametre olarak alıyor,gerekli ki-kare hesabını yapıyor ve F9 ile çalıştırıldığında sonuçları yazdırıyor. Toplam ki-kare hesabının yanında,her bir aralık için de ki-kare değerini ekstra olarak hesapladım.Son olarak tüm ki-kare değerlerini toplayarak ilgili eşik değeriyle kıyas yapıyor ve testin başarılı/başarısız olma durumunu ekrana yazdırıyor.

Hesaplamalar sonucunda ki-kare değerimi 139.8254 olarak buldum,bu ise ilgili eşik değerinin - 12.592- 10 katından daha fazla olduğundan,gözlem ve beklentimin ki-kare testine uymadığını gördüm.