

**TARAMA İSTATİSTİKLERİ VE TÜRKİYE DEPREM
VERİLERİNE UYGULAMASI**

**SCAN STATISTICS AND ITS APPLICATION TO TURKISH
EARTHQUAKE DATA**

RABİA BURCU DEMİRTÜRK

Hacettepe Üniversitesi

Lisansüstü Eğitim – Öğretim ve Sınav Yönetmeliğinin

İstatistik Anabilim Dalı İçin Öngördüğü

YÜKSEK LİSANS TEZİ

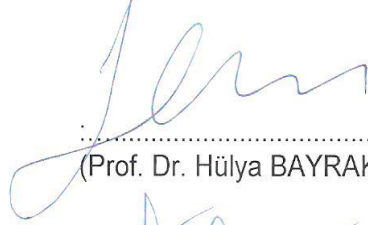
olarak hazırlanmıştır.

2012


Fen Bilimleri Enstitüsü Müdürlüğü'ne,

Bu çalışma jürimiz tarafından **İSTATİSTİK ANABİLİM DALI** 'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan


.....
(Prof. Dr. Hülya BAYRAK)

Üye


.....
(Prof. Dr. Gül ERGÜN)

Üye (Danışman)


.....
(Dr. Ayten YİĞİTER)

ONAY

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından/...../..... tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca/...../..... tarihinde kabul edilmiştir.

Prof.Dr. Fatma Sevin DÜZ
Fen Bilimleri Enstitüsü Müdürü

TARAMA İSTATİSTİKLERİ VE TÜRKİYE DEPREM VERİLERİNE UYGULAMASI

Rabia Burcu DEMİRTÜRK

ÖZ

Olaylar arasındaki anlamlı kümelenmelerin tespit edilmesinde kullanılan tarama istatistikleri uygulama alanı geniş olan bir yöntemdir. Bu çalışmanın amacı, tarama istatistiklerinin tanıtılması ve Türkiye deprem verileri üzerine bir uygulamasıdır. Bu amaçla, tek, iki, üç boyutlu tarama istatistiklerinin çeşitli olasılık dağılımları altındaki teorisi tez çalışmasında incelenmiştir. Tarama istatistiklerin elde edilmesinde kullanılan SaTScan yazılımı kısaca tanıtılmış ve bu yazılım yardımıyla Türkiye deprem verilerinin zaman, mekan, uzay-zaman boyutunda kümelenmeleri araştırılmıştır. Tez çalışması ülkemizde tarama istatistiklerinin kullanıldığı ilk çalışmadır. Elde edilen bulgular ülkemizde hassas bir konu olan deprem için faydalı bilgiler içermektedir.

Anahtar Kelimeler: Tarama İstatistiği, Mekansal İstatistik, Deprem Verileri

Danışman: Dr. Ayten YİĞİTER, Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Olasılık Teorisi ve Olasılık Süreçleri Anabilim Dalı

SCAN STATISTICS AND ITS APPLICATION TO TURKISH EARTHQUAKE DATA

Rabia Burcu DEMİRTÜRK

ABSTRACT

The scan statistics is a statistical method which detects the clusters among the events and has been widely used in various areas. The aim of the study is to introduce the scan statistics and apply that methodology to Turkish earthquake data. The theories of one, two, three dimensional scan statistics under the various probability distributions are investigated in the study. SaTScan software, which is used to obtain the scan statistics is briefly mentioned in the thesis and then the clusters in Turkish earthquake data are explored in temporal, spatial, space-time dimensional by using the software. The thesis contains the first application of the scan statistics in Turkey, especially it contains valuable results for the Turkish earthquake structure.

Key Words: Scan Statistic, Spatial Statistic, Earthquake Data

Advisor: Dr. Ayten YİĞİTER, Hacettepe University, Faculty of Science, Department of Statistics, Probability Theory and Stochastic Processes Section

TEŐEKKÜR

Çalıőmam boyunca bana her konuda destek olan, fikir ve önerileriyle yapabileceđimin en iyisi için beni heveslendiren, bilgi ve tecrübesi ile her zaman bana yol gösteren deđerli danıőmanım Sayın Dr. Ayten YİĐİTER'e,

Manevi desteđiyle beni hiçbir zaman yalnız bırakmayan, çalıőmamın tüm aőamalarında sabır ve anlayıő göstererek beni dinleyen, desteđini hep hissettiđim arkadaőım, meslektaőım Fatih DÜZGÜN'e,

Çalıőmam boyunca duygularıma ortak olan tüm dostlarıma,

Bu günlere gelebilmem için hiçbir fedakârlıktan kaçınmayan, her zaman yanımda olan canım annem Candan DEMİRTÜRK'e, canım babam Fatih DEMİRTÜRK'e ve çalıőmam boyunca onu ne kadar ihmal etsem de benden desteđini esirgemeyen canım kardeőim Aysu DEMİRTÜRK'e,

ÇOK TEŐEKKÜR EDERİM...

İÇİNDEKİLER DİZİNİ

ÖZ	iii
ABSTRACT.....	iv
TEŞEKKÜR	v
1. GİRİŞ.....	7
2. TARAMA İSTATİSTİKLERİ.....	9
2.1 Tek Boyutlu (Zamansal) Tarama İstatistiği	10
2.1.1 Geriye Doğru (Koşullu) Tarama.....	10
2.1.2 İleriye Doğru (Koşulsuz) Tarama.....	17
2.2 İki Boyutlu (Mekansal) Tarama İstatistiği.....	18
2.2.2 İki Boyutlu Tarama Penceresinin Şekli.....	36
2.3 Üç Boyutlu (Uzay-Zaman) Tarama İstatistiği.....	36
3. HİPOTEZ TESTİ ve TESTİN GÜCÜ	39
4. MONTE CARLO BENTEZİMİ ve ALGORİTMALAR	41
4.1. Algoritmalar	43
5. BAZI KULLANIM ALANLARI	46
6. SaTScan PROGRAMI	49
7. DEPREM OLGUSU ve BAZI TANIMLAR	52
8. UYGULAMALAR	56
8.1 TEK BOYUTLU TARAMA İSTATİSTİKLERİ.....	56
8.1.1 Aylık Deprem Verileri	56
8.1.2 Günlük Deprem Verileri	58
Çizelge 8.1. 2009-2010-2011 Deprem Verileri (Günlük) ve Geriye Doğru Kümelenmeler .	58
8.2 İKİ BOYUTLU TARAMA İSTATİSTİKLERİ	61
8.3. ÜÇ BOYUTLU TARAMA İSTATİSTİKLERİ	65
9. SONUÇLAR	70
KAYNAKLAR.....	72
EKLER.....	75
EK 1: Tek Boyutlu Aylık Deprem Verileri Geriye Doğru Tarama.....	75
EK 2: Tek Boyutlu Aylık 4'ten Büyük Deprem Verileri Geriye Doğru Tarama.....	75
EK 3: Tek Boyutlu Aylık Deprem Verileri İleriye Doğru Tarama.....	76
EK 4: Tek Boyutlu Aylık 4'ten Büyük Deprem Verileri İleriye Doğru Tarama.....	76
EK 5: Tek Boyutlu 2009 Yılı Günlük Deprem Verileri Geriye Doğru Tarama.....	77
EK 6: Tek Boyutlu 2010 Yılı Günlük Deprem Verileri Geriye Doğru Tarama.....	77
EK 7: Tek Boyutlu 2011 Yılı Günlük Deprem Verileri Geriye Doğru Tarama.....	77

EK 8: Tek Boyutlu 2009 Yılı Günlük Deprem Verileri İleriye Doğru Tarama	78
EK 9: Tek Boyutlu 2010 Yılı Günlük Deprem Verileri İleriye Doğru Tarama	78
EK 10: Tek Boyutlu 2011 Yılı Günlük Deprem Verileri İleriye Doğru Tarama.....	79
EK 11: İki Boyutlu 2009 Yılı Kümelenmeler.....	79
EK 12: İki Boyutlu 2010 Yılı Kümelenmeler.....	80
EK 13: İki Boyutlu 2011 Yılı Kümelenmeler.....	81
EK 14: Üç Boyutlu 2011 Yılı Kış Mevsimi Kümelenmeler	81
EK 15: Üç Boyutlu 2011 Yılı İlkbahar Mevsimi Kümelenmeler	82
EK 16: Üç Boyutlu 2011 Yılı Yaz Mevsimi Kümelenmeler	82
EK 17: Üç Boyutlu 2011 Yılı Sonbahar Mevsimi Kümelenmeler.....	83
ÖZGEÇMİŞ.....	84

ŞEKİLLER DİZİNİ

Şekil 2.1. 1991-1995 Yılları Arasında Görülen Kanser Vakaları (O:Ocak, Ş:Şubat,..., A:Aralık).....	11
Şekil 2.2. Birinci Durum ile İlgili Aralık ve Gözlem Sayısı	13
Şekil 2.3. $B * $ Olayının Gösterimi	13
Şekil 2.4. İkinci Durum ile İlgili Aralık	15
Şekil 2.5. $E(n N; u, v)$ Olayı	19

RESİMLER DİZİNİ

Resim 7.1. Türkiye Deprem Bölgeleri Haritası.....	52
Resim 8.1. Mekansal Kümelenmeler	64
Resim 8.2. 2011 Yılı Sonbahar Mevsimi En Olası Küme.....	68

ÇİZELGELER DİZİNİ

Çizelge 2.2. Katlı Terimli Dağılım Modeli İle İlgili Anlamlı Kümeler	35
Çizelge 8.1. 2009-2010-2011 Deprem Verileri (Günlük) ve Geriye Doğru Kümelenmeler	58
Çizelge 8.2. 2009-2010-2011 Deprem Verileri (Günlük) ve İleriye Doğru Kümelenmeler	60
Çizelge 8.3. Mekansal Tarama ile Bulunan En Olası Kümeler	62
Çizelge 8.4. Mekansal Tarama ile Bulunan İkinci Olası Kümeler.....	63
Çizelge 8.5. Uzay-Zaman Taraması ile Bulunan En Olası Kümeler	67

1. GİRİŞ

Günümüzde bilgisayar teknolojilerinin hızlı gelişmesi, istatistiksel çözümlerinin çeşitlenmesinde ve etkin kullanılmasında önemli bir etkidir. Bu hızlı gelişme, olaylar içindeki kümelenmelerin saptanmasında kullanılan bir yöntem olan tarama istatistiklerinin elde edilmesinde de önemli olmuştur. Birçok alanda araştırmacılar olayların kümelenmesine çok önem vermektedir. Bir epidemiyolojist kanser vakalarının coğrafi konumda kümelenmelerini belirlemeye çalışabilir, telekomünikasyon mühendisleri bir santrale eş zamanlı yapılan aramaların kümelenmelerine uyum sağlayacak şekilde bir kapasite tasarlaması yapabilirler. Kalite kontrol uzmanları üretim bandındaki kusurlu grupları araştırabilirler. Bir jeolog belirli türden maden kaynaklarının kümelendiği yerleri bulmak için belirlenen bölgeyi, bir astrofizikçi ise gama ışını yayan patlamaların yoğun olduğu kaynakları belirlemek için gökyüzünü tarayabilir. (Glaz et al., 2001).

Yukarıda bazı alanlarda örnekleri verilen araştırmalar tarama istatistiklerinin ne kadar yaygın kullanımı olduğunu göstermektedir.

Tarama İstatistikleri ilk olarak 1965 yılında Joseph Naus tarafından “Bir Çizgi Üzerindeki Noktaların Maksimum Küme Boyutunun Dağılımı” ve “İki Boyutlu Bir Düzlemde Rasgele Noktaların Kümelenmesi” çalışmalarıyla kullanılmaya başlamıştır. Daha sonra günümüze kadar pek çok alanda; tek, iki ya da daha çok boyutlu analizlerde Bernoulli, Poisson, normal, üstel vb. dağılıma sahip rasgele olaylar için tanımlanmıştır. Literatürde tek, iki, üç boyutlu tarama istatistiğinin özel bir durumu olan zamansal (temporal), mekansal (spatial), uzay-zaman (space-time) yönünden uygulamaları ile daha çok karşılaşılmaktadır.

Bu çalışmada tek, iki ve üç boyutlu tarama istatistiklerinin teorisi çeşitli dağılımlar altında incelenmiştir. Bu amaçla, çalışmanın Birinci Bölümü’nde konuya bir giriş yapılmış, İkinci bölüm’de ise, tarama istatistiklerinin tarihi gelişiminden bahsedildikten sonra tek, iki, üç boyutlu tarama istatistikleri ve uygulama alanları ile ilgili örnekler verilmiştir.

Üçüncü Bölüm’de, tarama istatistiklerinde amaçlanan anlamlı kümelerin olup olmadığı ile ilgili hipotez testine yer verilmiştir. Dördüncü Bölüm’de ise, tarama

istatistiklerinde hipotez testinde p deęerinin bulunmasında kullanılan Monte Carlo benzetimi ve tarama istatistiklerinin bulunması ile ilgili geliştirilen çeşitli algoritmalar verilmiştir.

Beşinci Bölüm'de, tarama istatistiklerinin kullanım alanları ve bazı örneklere yer verilmiştir.

Altıncı Bölüm'de, uygulamada kullanılan SaTScan programı tanıtılmıştır.

Yedinci Bölüm'de, çalışmanın uygulama aşamasında kullanılan Türkiye deprem verileri hakkında kısaca bilgi verilmiştir.

Sekizinci Bölüm'de ise, Türkiye'deki deprem verilerinin, tarama istatistikleri kullanılarak zamansal, mekansal, uzay-zaman yönünden kümelenmelerinin olup olmadığı araştırılmış ve çeşitli sonuçlar elde edilmiştir.

Sonuç bölümünü oluşturan Dokuzuncu Bölüm'de ise elde edilen bulgular yorumlanmıştır.

2. TARAMA İSTATİSTİKLERİ

Kümelenmenin tamamen rasgele mi oluştuğunu yoksa belirli bir olasılık modeline mi uygun olduğunu belirlemeyi amaçlayan tarama istatistikleri, olayların ileriye ve geriye doğru taranmasında kullanılabilir. Bu çalışmada tek boyutlu, iki boyutlu ve üç boyutlu tarama istatistiklerine yer verilmiş, literatürdeki uygulamaları üzerinde durulmuştur.

Tarama istatistiğinin babası Joseph Naus olarak bilinir. 1959 yılında New York City College'da işletme lisansını tamamlayan Joseph Naus bir sonraki yıl Harvard'da ekonomi üzerine yüksek lisans yapmaya başlamıştır. O sırada danışmanı olan Robert Dorfman, mezuniyet çalışmaları için bilgi düzeyini arttırması bakımından ekonomi alanını istatistik gibi çeşitli alanlarla birleştirmesini önermiştir. İlk istatistik derslerini Arthur Dempster'den alan Naus istatistiğin ekonomiye göre daha geniş kullanım alanı olduğunu ve daha kullanışlı olduğunu görmüş, böylece ilk yılında istatistik bölümüne geçiş yapmıştır. Naus, istatistik bölümündeki ikinci yılında zamanının oldukça büyük bir kısmını Deniz Kuvvetleri tarafından finanse edilen Operasyon Değerlendirme Grubu Fen Bilimleri Bölümü'nde (Applied Science Division of the Operations Evaluation Group) geçirdi. Buradaki ilk problem donanma ihtiyaçlarından kaynaklanmıştır. Deniz Kuvvetleri donanmanın okyanusun bir bölgesindeki yığılmalarının araştırılmasını istemiştir. Bununla birlikte buradaki problem iki boyutlu olarak görünse de daha genel koşullar içinde yer almıştır. Operasyon Değerlendirme Grubunun bir üyesi olan Naus, 1961 yılının sonbaharında çalışmasına başlamış ve 1962 yılında sonuçlarını Fen Bilimleri Bölümü'nde yayınlamıştır. Tezinden önce yayınlanan bu teknik rapor birçok araştırmacıya referans olmuştur. Bu araştırma doğrultusunda Deniz Kuvvetleri'yle tekrar bir sözleşme yapılarak araştırmalara devam edilmiştir. Naus, Ekim 1963'te Harvard Üniversitesi İstatistik Bölümü'nde "Bir Doğru ve Düzlem Üzerindeki Rasgele Noktaların Kümelenmesi" başlıklı tezini sunmuş ve tezi onaylanmıştır. Bir sonraki yılın Ocak ayında ise kendisine doktora derecesi verilmiştir. 1963-1964 yılları arasında araştırma analisti olarak çalışmalarına Deniz Araştırmaları Enstitüsü'nde devam etmiştir. Harvard'dan sonra 1964-1966 yıllarında Rutgers Üniversitesi'nde istatistik alanında yardımcı doçent olarak göreve başlamıştır. Bu dönemde Naus ve diğer öğretim üyeleri arasındaki işbirliği ve uyum çok verimli olmuştur. 1965 yılında

“Bir Doğru Üzerindeki Noktaların Maksimum Küme Boyutunun Dağılımı” adlı ilk makalesini yayınlarken aynı yıl tarama istatistikleri üzerinde çalışmaya başlamasının nedeni olan iki boyutlu problem ile ilgili ilk makalesini de yayınlamıştır.

Naus’un çalışmalarıyla başlayan tarama istatistiklerinin tek, iki, üç boyutlu modelleri bundan sonraki kesimlerde anlatılmıştır.

2.1 Tek Boyutlu (Zamansal) Tarama İstatistiği

Tek boyutlu tarama istatistiği, 1965 yılında Naus tarafından bir doğru üzerindeki noktaların maksimum küme boyutunun dağılımının bulunması ile ilgili çalışmasından sonra zaman içinde çeşitli yönlerde genişletilmiştir. Örneğin, Saperstein (1972) ve Naus (1974) Bernoulli modeli ile ilgili, Weinstock (1981) kümelenmelerin tespitinde genelleştirilmiş tarama istatistiği ile ilgili çalışma yapmışlardır. Ayrıca Glaz and Naus (1983) tarama istatistiğini bir doğru üzerinde ortaya çıkan olayların birden çok kümelenmelerinin araştırılmasında kullanmışlardır. Wallenstein et al. (1989) Amerika’daki gençlerin işlediği suçlara ilişkin mevsimsel kümelenmelerin araştırılması üzerine bir çalışma yapmışlardır. Loader (1991) ise, tek boyutlu tarama istatistiklerinde sabit olmayan pencere büyüklüğünü araştırmıştır.

2.1.1 Geriye Doğru (Koşullu) Tarama

Bir aralıkta N tane olayın (noktanın) verilmesi koşulu altında olayların kümelenmesine ilişkin olasılıkların hesaplanmasında tanımlanan tarama işlemi geriye doğru tarama olarak adlandırılır.

N tane olay, $(0,1)$ gibi bir aralıkta ya da zaman aralığında $x_1 \leq x_2 \leq \dots \leq x_N$ şeklinde rasgele sıralansın. Bu olaylardan n tanesinin herhangi bir w uzunlukta bir aralıkta görülmesi olasılığı sıralı istatistiklerin teorisinden yararlanılarak bulunur. Bu olasılık,

$$\begin{aligned} b(n|N; w) &= \binom{N}{n} \left(\int_0^w dx \right)^n \left(\int_w^1 dx \right)^{N-n} \\ &= \binom{N}{n} w^n (1-w)^{N-n} \end{aligned} \quad (2.1)$$

ile hesaplanır. Benzer şekilde bu olaylardan en çok n tanesinin w uzunluklu bir aralıkta görülmesi olasılığı,

$$F_b(n|N; w) = \sum_{i=0}^n b(i|N; w) \quad (2.2)$$

ile, en az n tanesinin w uzunluklu bir aralıkta görülmesi olasılığı ise,

$$G_b(n|N; w) = \sum_{i=n}^N b(i|N; w) \quad (2.3)$$

ile hesaplanmaktadır.

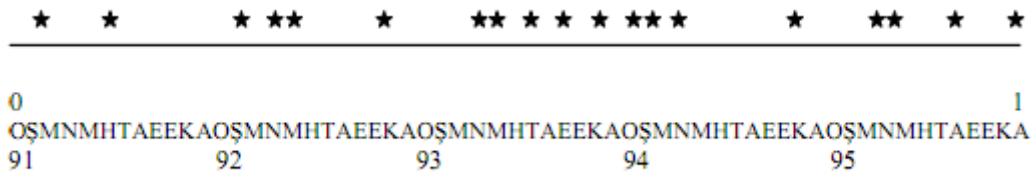
w uzunluğundaki sabit bir aralıkta ya da zaman aralığında gerçekleşen olayların maksimum sayısı S_w olsun. Bu rasgele değişkene tarama istatistiği (scan statistic) adı verilir. Bir diğer rasgele değişken W_n , verilen sabit n sayıda olay içeren en kısa zaman aralığıdır. S_w ve W_n istatistiklerinin dağılımları arasında,

$$P(W_n \geq w) = P(S_w \leq n) \quad (2.4)$$

eşitliği vardır. Bu eşitlik, n sayıda olay içeren en kısa aralığın en az w uzunluğunda olması olasılığı, maksimum olay sayısının en fazla n olması olasılığına eşittir şeklinde yorumlanabilir (Glaz et al., 2001).

Tarama istatistiğinin gerekliliği aşağıda verilen bir örnekle açıklanabilir:

Bir şehirde 1991–1995 yılları arasındaki 5 yıllık bir dönemde 19 adet belirli bir türde kanser vakası tespit edilmiştir. Veriler incelendiğinde 14 Nisan 1993 ile 13 Nisan 1994 tarihlerini kapsayan 1 yıllık dönemde 8 olayın gerçekleştiği gözlenmiştir.



Şekil 2.1. 1991-1995 Yılları Arasında Görülen Kanser Vakaları (O:Ocak, Ş:Şubat,..., A:Aralık)

Bu şekilde 1 Ocak 1991 ile 31 Aralık 1995 tarihleri (0,1) birim zaman aralığı olarak gösterilmiş ve bu dönem $w = 0,2$ uzunluklu bir aralıkla taranmıştır. Burada, $N = 19$, $S_{0,2} = 8$ olmaktadır. Bu tarama aralığında gözlenen maksimum kanser vakası 8'dir.

Bu örnek için, 19 olayın gerçekleştiği 5 yıllık bir dönemde, herhangi bir 1 yıllık zaman aralığının 8 ya da daha çok olay içermesi olağan mıdır (tekbiçimli dağılıma uygun mudur) sorusu sorulabilir. 19 olayın her biri öteki olaylardan bağımsız olarak bu 1 yıllık zaman aralığının içinde gözlenebilirdi. O halde 8 veya daha fazla olayın bu 1 yıllık zaman aralığında gözlenmesi olasılığı, $N = 19; p = \frac{1}{5}$ alınarak iki terimli (binom) dağılımı yardımıyla hesaplanabilirdi. Fakat bu hesaplama bize 14 Nisan 1993 ile 13 Nisan 1994 tarihleri ile tanımlanan özel bir yıl değil, herhangi 1 yıllık zaman aralığının 8 veya daha fazla olay içermesinin tekbiçimli dağılıma uygun olup olmadığı sorusuna cevap vermez.

Ayrıca bu problem, 5 yıllık dönem ayırık 1 yıllık zaman aralıklarına bölünerek ve herhangi bir yılda gözlenen maksimum olay sayısının dağılımı kullanılarak da çözülmeye çalışılabilir. Yine de bu sorunun cevabı değildir. Çünkü maksimum olay sayısının gözlendiği yıl ayrılmış iki yılın arasında kalmış olarak gözlenebilir (Glaz et al., 2001).

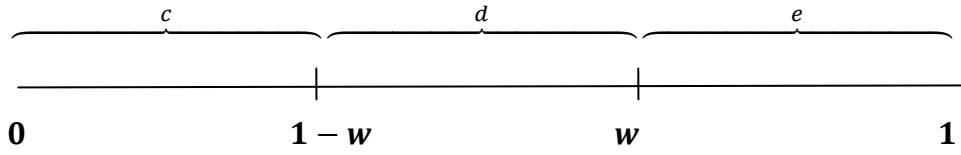
Yukarıda verilen nedenlerden dolayı bu tür bir sorunun çözümünde tek boyutlu tarama istatistikleri kullanılmaktadır.

(0,1) aralığında rasgele sıralanan N tane olaydan n tanesinin w uzunluğundaki tarama penceresinde ortaya çıkma olasılığı aşağıdaki gibidir:

$$\begin{aligned}
 P(n|N; w) &= P(A^*) + P(B^*), \quad w \geq \frac{1}{2} \text{ ve } n > \frac{N+1}{2} \\
 &= P(A) + P(B), \quad w \leq \frac{1}{2} \text{ ve } n > \frac{N}{2}.
 \end{aligned} \tag{2.5}$$

Eşitlik (2.5)'te verilen olasılıktaki 1.durum, $w \geq \frac{1}{2}$ ve $n > \frac{N+1}{2}$ 'yi ele alalım:

(0,1) zaman aralığında N tane olay rasgele sıralansın. c , $(0, 1-w)$ aralığında gözlenen olay sayısını, d , $(1-w, w)$ aralığında gözlenen olay sayısını ve e , $(w, 1)$ aralığında gözlenen olay sayısını göstereyin ($c + d + e = N$).



Şekil 2.2. Birinci Durum ile İlgili Aralık ve Gözlem Sayısı

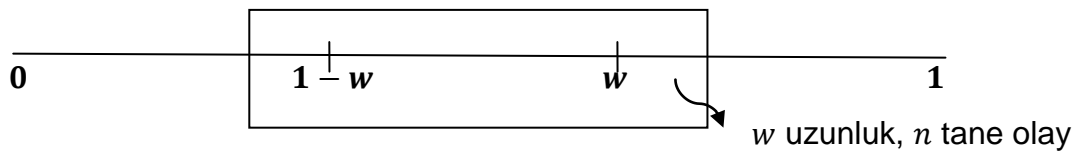
Verilen aralıklardaki olay sayılarının birlikte ortaya çıkma olasılığı aşağıdaki gibi bulunur:

$$\begin{aligned}
 P(c, d, e) &= \frac{N!}{c!d!e!} \left(\int_0^{1-w} dx \right)^c \left(\int_{1-w}^w dx \right)^d \left(\int_w^1 dx \right)^e \\
 &= \frac{N!}{c!d!e!} (1-w)^c (2w-1)^d (1-w)^e
 \end{aligned} \tag{2.6}$$

Bu durumda, bir w uzunluğunda n tane olayın çıkması olasılığı $P(n|N; w)$, ayrık A^* ve B^* olaylarının olasılıklarının toplamına eşittir. Burada,

A^* olayı, $c + d \geq n$ olması ya da $e + d \geq n$ olması ya da her iki durumun da meydana gelmesi olayını gösterebilirsin,

B^* olayı, $c + d < n$ ve $e + d < n$ olaylarının birlikte meydana gelmesi ve $(0, 1-w)$ aralığı ile $(w, 1)$ aralığının kesişiminin w uzunluklu olması ve n sayıda olay içermesini gösterebilirsin.



Şekil 2.3. B^* Olayının Gösterimi

A^* olayının olasılığı,

$$P(A^*) = P(c + d \geq n) + P(d + e \geq n) - P(c + d \geq n \cap d + e \geq n)$$

olur. Burada,

$$P(c + d \geq n) = \sum_{i=n}^N \binom{N}{i} w^i (1-w)^{N-i} = G_b(n|N; w),$$

$$P(d + e \geq n) = \sum_{i=n}^N \binom{N}{i} w^i (1-w)^{N-i} = G_b(n|N; w),$$

$$\begin{aligned} P(c + d \geq n \cap d + e \geq n) &= \underbrace{P(c + d = n)P(d + e \geq n|c + d = n)}_{y=n} \\ &+ \underbrace{P(c + d = n + 1)P(d + e \geq n + 1|c + d = n + 1)}_{y=n+1} \\ &+ \dots + \underbrace{P(c + d = N)P(d + e = N|c + d = N)}_{y=N} \end{aligned}$$

biçimindedir. Böylece,

$$P(A^*) = 2G_b(n|N; w) - \sum_{y=n}^N b(y|N; w)F_b\left(N - n|y; \frac{1-w}{w}\right)$$

olarak bulunur.

B^* olayının meydana gelmesi olasılığı hesaplanırken, w uzunluğu $(0,1)$ uzunluklu aralık boyunca hareket ettirilir. Bu durumda, w uzunluğu her zaman $(1-w, w)$ aralığını kapsar. Bu nedenle d tane olay her zaman bu aralıkta bulunur. w uzunluklu aralık hareket ettirildiği sürede c olay bu aralıktan çıkarken, e tane olay bu aralık içine girmektedir. Böylece bir yol tanımlanmış olur. Bu yol $c + d$ ile başlar $d + e$ ile biter. Bunun sonucunda da tüm olası durumlara göre, B^* olayının koşullu olasılığı aşağıdaki gibi hesaplanır:

$$P(B^*|c, d, e) = \frac{\binom{c+e}{n-d} \binom{N}{d}}{\frac{N!}{(c! d! e!)}}$$

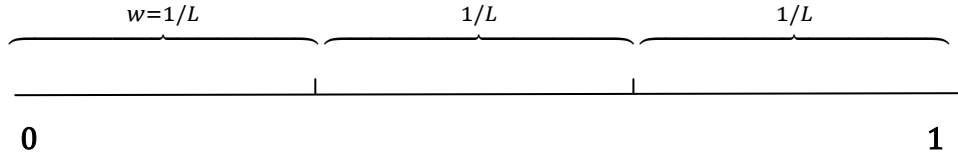
Buna göre, B^* olayının olasılığı, Eşitlik (2.6)'da verilen olasılık kullanılarak ve c, d, e 'nin alacağı olası değerler üzerinden toplanarak aşağıdaki gibi bulunur:

$$P(B^*) = \sum_{d=0}^{2n-N-2} \sum_{c=N-n+1}^{n-d-1} \binom{N}{n} \binom{n}{d} (2w-1)^d (1-w)^{N-d}.$$

Eşitlik (2.5)'teki 1. durumdaki olasılık A^* ve B^* olaylarının olasılıkları toplamına eşit olur.

Eşitlik (2.5)'te verilen olasılıktaki 2.durum, $w \leq \frac{1}{2}$ ve $n > \frac{N}{2}$ yi ele alalım:

Bu durumda $w = \frac{1}{L}$ olarak tanımlanır. L sabit bir sayıdır. $(0,1)$ aralığı L tane aralığa bölünür.



Şekil 2.4. İkinci Durum ile İlgili Aralık

Burada yine,

$$P(n|N; w) = P(A) + P(B)$$

eşitliği vardır. Burada, A olayı, en az bir aralığın en az n tane olay içermesini gösterir ve A olayının olasılığı,

$$P(A) = L \sum_{i=n}^N b(i|N; w) = LG_b(n|N; w)$$

biçimindedir. B_i olayı, $(0,1)$ aralığının arakesiti olmayan L tane aralığın en çok n olay içermesini ve i . ile $i + 1$. aralıkların kesişiminin w uzunluklu olup en az n tane olay içermesini gösterebilir. B olayı bu tanımlanan B_i olaylarının birleşimidir:

$$P(B) = \sum_{i=1}^{L-1} P(B_i) - \sum_{i=1}^{L-2} P(B_i \cap B_{i+1}).$$

B_i olasılıkları hesaplanırken, c , i . aralıktaki olayların sayısı, d , $i + 1$. aralıktaki olayların sayısı ve e , geri kalan olayların sayısı olarak tanımlansın ($c + d + e = N$).

Verilen aralıklardaki c, d, e sayıda olayın birlikte ortaya çıkma olasılığı,

$$P(c, d, e) = \frac{N!}{c! d! e!} \left(\int_0^w dx \right)^c \left(\int_w^{2w} dx \right)^d \left(\int_{2w}^1 dx \right)^e$$

$$= \frac{N!}{c!d!e!} w^c w^d (-2w)^e \quad (2.7)$$

ile bulunur.

Bir w aralığı $(i - 1)w$ ile $(i + 1)w$ aralıkları arasında hareket ettirildiğinde meydana gelen yol c ile başlar ve d ile biter. Bu durumda c azaldığında d artar. Buna göre koşullu B_i olasılıkları şu şekilde hesaplanmaktadır:

$$P(B_i|c, d, e) = \frac{\binom{c+d}{n} \binom{N}{e}}{\frac{N!}{c!d!e!}}. \quad (2.8)$$

Eşitlik (2.8)'den B_i olayının olasılığı, Eşitlik (2.7)'de verilen olasılıktan yararlanarak ve c, d, e 'nin alacağı olası değerler üzerinden toplanarak aşağıdaki gibi bulunur:

$$P(B_i) = \sum_{e=0}^{N-n} \sum_{c=N-n-e+1}^{n-1} \binom{N}{n} \binom{N-n}{e} (1-2w)^e w^{N-e}.$$

i . ile $i + 1$. aralıkların kesişiminin olasılıkları ise,

$$P(B_i \cap B_{i+1}) = G_b(n + 1|N; w)$$

biçiminde hesaplanır. Eşitlik (2.5)'teki 2. durumdaki olasılık A ve B olaylarının olasılıkları toplamına eşit olur (Naus, 1965a).

Geriye doğru tarama olasılıklarının hesaplanması ile ilgili aşağıdaki örnek verilebilir:

1 dakikalık süre içerisinde rasgele zamanlarda yapılacak olan 15 telefon görüşmesi için aramalar başlatılmıştır. Bir görüşme için arama süresi 10 saniyedir. 8 ya da daha çok telefon görüşmesi için aynı anda arama yapılması olasılığının bulunması ile ilgilenelim.

$i = 1, 2, \dots, 8$ olmak üzere $(7 + i)$. telefon görüşmesi için i . görüşme başladıktan 10 saniye sonra arama yapılıyor. Bu durumda olasılık Eşitlik (2.5)'ten,

$$P\left(8|15; \frac{1}{6}\right) = 0,037$$

olarak hesaplanır.

Bu olasılık hesabı, 6 ayırık aralığın her birinin en az 8 aramayı kapsamaması olasılığından farklıdır. Bu son olasılık ise Eşitlik (2.3)'ten,

$$6G_b \left(8|15; \frac{1}{6} \right) = 0,008$$

olarak hesaplanır (Naus, 1965a).

2.1.2 İleriye Doğru (Koşulsuz) Tarama

Tek boyutlu tarama istatistiklerinde, belli bir zaman aralığında meydana gelmiş sabit N sayıda olaya bağlı olarak sonuç çıkarma işlemi geriye doğru tarama olarak tanımlanırken, bilinen bir olasılık dağılımına sahip N sayıda olayda olay sayısının da rasgele değişken olduğu durumlarda yapılan tarama ise ileriye doğru tarama olarak adlandırılmaktadır.

İleriye doğru tarama zaman içinde gözlenen olaylar için kullanılmaktadır. Örneğin, olayların Poisson sürecine uygun olarak ortaya çıktıklarını varsayalım. Poisson sürecinde belli bir zaman aralığında meydana gelen olay sayısı rasgele değişken ve birim zamanda ortaya çıkması beklenen olay sayısı λ 'dır. Örneğin bu sürecin kullanım alanları arasında telefon trafiği, kuyruk modelleri gösterilebilir.

Herhangi bir $[t, t + w]$ zaman aralığında meydana gelen olay sayısı $N_t(w)$, Poisson dağılımlı olup beklenen değeri λw 'dur. Bu durumda herhangi bir zaman aralığında meydana gelen olay sayısının n 'e eşit olması olasılığı aşağıdaki gibi hesaplanır:

$$P(N_t(w) = n) = \frac{e^{-\lambda w} (\lambda w)^n}{n!}, \quad n = 0, 1, 2, \dots$$

Her bir ara kesiti olmayan zaman aralığında meydana gelen olay sayısı birbirinden bağımsız olarak Poisson dağılımına uygun dağılmaktadır. Bir zaman aralığında meydana gelen olaylar için $T_{n,w}$, w uzunluklu bir aralıkta en az n tane olayın ilk kez görülmesine kadar geçen süredir. En küçük i değeri için $X_{(i+n-1)} - X_i \leq w$ olduğunda $T_{n,w} = X_{(i+n-1)}$ olmaktadır. $T_{n,w}$, S_w ve W_n istatistikleri arasındaki ilişki,

$$P(S_w \geq n) = P(W_n \leq w) = P(T_{n,w} \leq T)$$

şeklinde ifade edilmektedir. Poisson sürecinde $(0, T)$ zaman aralığında ortaya çıkan olayların sayısının dağılımı λT ile Poisson dağılımlıdır. $\frac{w}{T}$ zaman aralığında en az n tane olay çıkması olasılığı,

$$P\left(n; \lambda T, \frac{w}{T}\right) \approx 1 - \exp\left\{\frac{-\lambda^n w^{n-1} T}{(n-1)!}\right\} \quad (2.9)$$

ile bulunur.

İleriye dönük tarama olasılıklarının hesaplanması ile ilgili aşağıdaki örnek verilebilir:

Bir ilde 8 yıllık bir süre içerisinde birbirinden farklı zaman ve konumlarda 19 karbon monoksit zehirlenmesi meydana gelmiştir. Bir halk sağlığı çalışanı bu kadar zehirlenmenin sıra dışı olduğunu düşünerek gelecekteki bir 8 yıl içerisinde 8 ya da daha fazla olayı kapsayan bir 1 yıllık dönemin olmasının ne kadar mümkün olabileceğini araştırmak istemiştir.

Eşitlik (2.9)'a göre yapılan hesaplama sonucunda,

$$P\left(8; 19, \frac{1}{8}\right) = 0,09293$$

olarak bulunmuştur. Anlamlılık düzeyi 0,05 olarak alındığında bu kümelenmenin sıradan bir olay olacağı görülmektedir (Glaz et al., 2001).

2.2 İki Boyutlu (Mekansal) Tarama İstatistiği

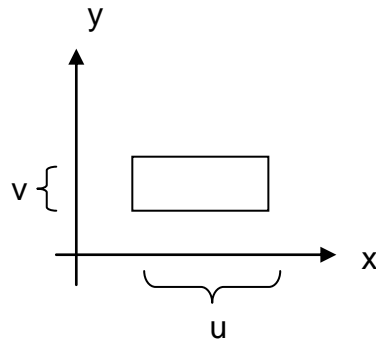
Tek boyutlu tarama istatistiklerinde, S_w 'nin tam olarak dağılımı Kesim 2.1'de açıklanan belli özel durumlarda bulunabilir. Daha fazla boyutlu taramalarda ise tarama istatistiğinin olasılık dağılımını bulmak daha karmaşıktır. Naus (1965b) "İki Boyutlu Bir Düzlemde Rasgele Noktaların Kümelenmesi" başlıklı çalışmasında sabit fakat keyfi bir büyüklükte dikdörtgen tarama penceresi kullanmıştır. Turnbull et al. (1990) hastalık kümelerinin bulunmasında dairesel tarama pencere kullanmışlar ve test istatistiğinin dağılımı tam olarak bulunamadığından Monte Carlo benzetiminden yararlanmışlardır.

İki boyutlu taramalarda tarama penceresi olarak kare, dikdörtgen, daire, üçgen vb. şekiller kullanılabilir. Taranan iki boyutlu bölgeler, dikdörtgensel bölgeleri, bir

küre yüzeyini ya da daha genel olarak belirli bir şekli olmayan coğrafik alanları içerir. Bu tür tarama istatistiklerinin dağılımları taranan bölgelerin yapısına göre elde edilmiştir. Sürekli yapıya sahip tarama bölgeleri için tekbiçimli, Poisson gibi dağılımlar, kafes (lattice) yapıda olanlar için iki terimli, hipergeometrik gibi dağılımlar kullanılır.

İki boyutlu tarama istatistiklerinde, tarama istatistiği S_w , şekli verilen w boyutlarında herhangi bir pencerede bulunan noktaların maksimum sayısıdır. W_n , n nokta içeren iki boyutlu en küçük tarama penceresinin boyutlarını gösterir (Glaz et al., 2001).

İki boyutlu tarama olasılıklarının bulunması ile ilgili birim karede rasgele dağılmış N olayın x_1, x_2, \dots, x_N olduğu durumu ele alalım. $E(n|N; u, v)$ olayı, birim kare içinde, yüksekliği u , genişliği v olan bir dikdörtgenin en az n tane olay içermesini gösterir ($0 \leq u, v \leq 1$).



Şekil 2.5. $E(n|N; u, v)$ Olayı

$P(n|N; u, v)$ olasılığı $E(n|N; u, v)$ olayının ortaya çıkma olasılığıdır. Burada,

$$E(n|N; u, 1) = E(n|N; u),$$

$$E(n|N; 1, v) = E(n|N; v)$$

olayları iki boyutlu tarama istatistiğinin tek boyuta indirgenmiş hali olup bu olayların olasılıkları ise sırasıyla,

$$P(n|N; u, 1) = P(n|N; u),$$

$$P(n|N; 1, v) = P(n|N; v)$$

şeklinde ifade edilir. $E(n|N; u, v)$ olayı $E(n|N; u)$ ve $E(n|N; v)$ olaylarının kesişimini ifade eder.

$n \leq N$ için,

$$\max \left\{ \begin{array}{c} P(n|N; u)P(n|n; v) \\ ; \\ P(n|N; v)P(n|n; u) \end{array} \right\} \leq P(n|N; u, v) \leq P(n|N; u)P(n|N; v) \quad (2.10)$$

eşitsizliği elde edilir.

$n = N$ olduğu durumda Eşitlik (2.10)'daki eşitsizliğin alt ve üst sınırları eşitlenir. Eşitlik (2.10)'da verilen ifadenin üst sınırı, $P(N|N; u) = Nu^{N-1} - (N-1)u^N$ olasılığından yararlanılarak,

$$\begin{aligned} P(N|N; u, v) &= P(N|N; u)P(N|N; v) \\ &= (Nu^{N-1} - (N-1)u^N)(Nv^{N-1} - (N-1)v^N) \\ &= N^2(uv)^{N-1} - Nu^{N-1}(N-1)v^N - (N-1)u^N Nv^{N-1} + (N-1)^2(uv)^N \\ &= N^2(uv)^{N-1} + (N-1)^2(uv)^N - N(N-1)(uv)^{N-1}(u+v) \end{aligned} \quad (2.11)$$

biçiminde bulunur (Burnside, 1936).

İki boyutlu tarama olasılıklarının tarama penceresinin biçimine göre hesaplanması ile ilgili aşağıdaki örnekler verilebilir:

Örnek1: Bir araştırmacı, kare ve dikdörtgen tarama pencerelerinden hangisinin bir kümenin görülme olasılığını daha yüksek hesaplayacağını araştırmak istiyor. Bu olasılık dikdörtgen tarama penceresine göre Eşitlik (2.11)'den,

$$P(5|5; 0,1; 0,9) = 0,00042$$

olarak hesaplanırken, kare tarama penceresine göre,

$$P(5|5; 0,3; 0,3) = 0,00095$$

olarak bulunmuştur. Yapılan hesaplamalar sonucunda, kare tarama penceresinde olasılık daha yüksek hesaplanmıştır. Çünkü $P(N|N; u, v)$ olasılığını maksimum yapabilmek için $u + v$ toplamının minimum olması gerekmektedir.

Örnek 2: Beş gemi okyanusta 10° enlem ve 10° boylam üzerindeki bir kare içinde birbirlerinden bağımsız rasgele olarak dağılmıştır. En az 4 geminin 2° enlem ve 3° boylam üzerinde olması olasılığı Eşitlik (2.10)'dan,

$$\max \left\{ \begin{array}{l} P(4|5; 0,2)P(4|4; 0,3) \\ ; \\ P(4|5; 0,3)P(4|4; 0,2) \end{array} \right\} \leq P(4|5; 0,2; 0,3) \leq P(4|5; 0,2)P(4|5; 0,3),$$

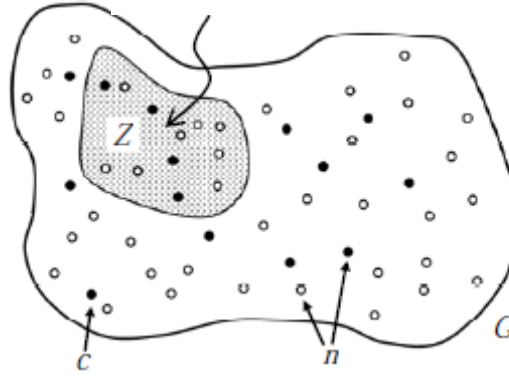
$$0,0086 \leq P(4|5; 0,2; 0,3) \leq 0,0278$$

olarak bulunur (Naus, 1965b).

Bernoulli, Poisson, normal vb. dağılımlı olayların iki boyutlu kümelenmesinde tanımlanan tarama istatistikleri ve hipotez testleri bunu izleyen kesimlerde açıklanacaktır.

2.2.1 Bernoulli Modeli

$p = \frac{c(Z)}{n(Z)}$ oranının yüksek olup olmadığı araştırılır.



Şekil 2.2. Bernoulli Dağılımına Sahip Olayların Mekansal Dağılımı

G ile gösterilen bir çalışma alanında, Z bir alt bölge olarak tanımlanmaktadır. Belirli özellikteki olayların bu alt bölge içerisinde olması olasılığı p iken, alt bölge dışında olması olasılığı ise q ile gösterilmektedir. Her bir olayın Z bölgesi içinde bulunması olasılığı birbirinden bağımsızdır. Burada,

$n(G)$: çalışma alanındaki toplam gözlem sayısı,

$n(Z)$: Z bölgesindeki toplam gözlem sayısı,

$c(G)$: çalışma alanındaki belirli özellik gösteren gözlem sayısı,

$c(Z)$: Z bölgesindeki belirli özellik gösteren gözlem sayısı

olarak tanımlansın. Bernoulli modeli için kümelenmenin olup olmadığına ilişkin hipotezler,

$$H_0: p = q,$$

$$H_1: p > q$$

biçimindedir.

Bu hipotezler aşağıdaki gibi de tanımlanabilir:

$A \subset G$ olacak şekilde bir bölge olsun.

$$H_0: Y_A \sim \text{Binom}(n(A), p) \quad \text{tüm } A \text{ alt bölgeleri için,}$$

$$H_1: Y_A \sim \text{Binom}(n(A), p) \quad A \subset Z,$$

$$: Y_A \sim \text{Binom}(n(A), q) \quad A \subset Z'$$

G alanında belirli özellik gösteren olayların Z alt bölgesinde olması olasılığı,

$$f(c(G), p) = \begin{cases} p, & c(G) \in Z \\ q, & c(G) \notin Z \\ 0, & \text{ö. d.} \end{cases}$$

ile hesaplanır. Seçenek hipotez altındaki olabilirlik fonksiyonu,

$$L(Z, p, q) = p^{c(Z)} q^{n(Z)-c(Z)} q^{c(G)-c(Z)} p^{(n(G)-n(Z))-(c(G)-c(Z))} \quad (2.12)$$

biçimindedir. Bilinmeyen parametreler p ve q 'nin en çok olabilirlik tahminleri,

$$\hat{p} = \frac{c(Z)}{n(Z)} \quad \text{ve} \quad \hat{q} = \frac{c(G)-c(Z)}{n(G)-n(Z)} \quad \text{Eşitlik (2.12)'de yerine konursa,}$$

$$L(Z) \stackrel{\text{def}}{=} \sup_{p>q} L(Z, p, q) = \left(\frac{c(Z)}{n(Z)} \right)^{c(Z)} \left(1 - \frac{c(Z)}{n(Z)} \right)^{n(Z)-c(Z)} \\ \times \left(\frac{c(G)-c(Z)}{n(G)-n(Z)} \right)^{c(G)-c(Z)} \left(1 - \frac{c(G)-c(Z)}{n(G)-n(Z)} \right)^{(n(G)-n(Z))-(c(G)-c(Z))}$$

elde edilir.

$$\frac{c(Z)}{n(Z)} > \frac{c(G)-c(Z)}{n(G)-n(Z)} \text{ iken,}$$

$$L(Z) = \left(\frac{c(G)}{n(G)} \right)^{c(G)} \left(\frac{n(G)-c(G)}{n(G)} \right)^{n(G)-c(G)}$$

olarak bulunur.

Yokluk hipotezi altındaki olabilirlik fonksiyonu aşağıdaki gibidir:

$$L_0 \stackrel{\text{def}}{=} \sup_{p=q} L(Z, p, q) = \left(\frac{c(G)}{n(G)} \right)^{c(G)} \left(\frac{n(G)-c(G)}{n(G)} \right)^{n(G)-c(G)} .$$

Böylece olabilirlik oranı λ ,

$$\lambda = \frac{\max_z L(Z)}{L_0}$$

biçiminde tanımlanır. Ayrıca olabilirlik oranı λ tarama istatistiği olarak da tanımlanır. Olabilirlik oranı, olası tüm alt bölgeler üzerinden maksimum yapılmaya çalışılır. Olabilirlik oranını maksimum yapan Z bölgesi en olası kümeyi gösterir.

Bernoulli modeli altında iki boyutlu tarama istatistiğine ilişkin bir örnek aşağıda verilmektedir:

Kuzey Carolina'da ani bebek ölümü üzerine 1974-1984 yılları arasında 100 ilin her birindeki canlı doğum ile ani bebek ölümleri ile ilgili kayıtlar kullanılarak tarama istatistiklerinde Bernoulli modeli ile analiz yapılmış ve çeşitli sonuçlar elde edilmiştir.

İllerdeki toplam canlı doğum sayısı $n(G)$, 753354 olup, toplam ani bebek ölüm sayısı $c(G)$, 1503'tür. İl merkezlerinin konumu coğrafi koordinat olarak kullanılmıştır (Kulldorff, 1997).

Değişen çaplı dairesel tarama penceresi kullanılarak yapılan analiz sonucunda bulunan anlamlı kümeler ve özellikleri aşağıdaki gibidir:

Bölge (Z)	Ani Bebek Ölüm Sayısı (c(Z))	Canlı Doğum Sayısı (n(Z))	p Değeri
A	139	36376	0,0001
B	59	14388	0,0005

Çizelge 2.1. Bernoulli Modeli Örneği için Bulunan Anlamlı Kümeler

2.2.2 Poisson Modeli

Belirli özellikteki olayların bir alt bölge içerisinde olması olasılığı p iken, bu alt bölge dışında olması olasılığı ise q ile gösterilmektedir.

$A \subset G$ olacak şekilde bir bölge olmak üzere, A bölgesinde ortaya çıkan olay sayısı Y_A 'nın dağılımı,

$$Y_A \sim \text{Poisson}(pn(A \cap Z) + qn(A \cap Z'))$$

biçimindedir. Poisson modeli için kümelenmenin olup olmadığına ilişkin hipotezler,

$$H_0: Y_A \sim \text{Poisson}(pn(A)) \quad \text{tüm } A \text{ alt bölgeleri için,}$$

$$H_1: Y_A \sim \text{Poisson}(pn(A)) \quad A \subset Z,$$

$$: Y_A \sim \text{Poisson}(qn(A)) \quad A \subset Z'$$

şeklinde kurulur.

G alanında belirli özellik gösteren olayların sayısının olasılık fonksiyonu,

$$f(c(G)) = \frac{\exp[-pn(Z) - q(n(G) - n(Z))][pn(Z) + q(n(G) - n(Z))]^{c(G)}}{c(G)!}, \quad c(G) = 0,1,2 \dots$$

$$= 0, \quad \text{ö. d.} \quad (2.13)$$

biçimindedir. Herhangi bir x konumunda gözlenen olayın Z bölgesine göre olasılık fonksiyonu (insidansı),

$$\begin{cases} \frac{pn(x)}{pn(Z)+q(n(G)-n(Z))}, x \in Z \\ \frac{qn(x)}{qn(Z)+q(n(G)-n(Z))}, x \notin Z \end{cases} \quad (2.14)$$

olarak tanımlanır. Seçenek hipotez altında x_i konumlarına göre belirli özellikteki noktaların sayısına ilişkin olabilirlik fonksiyonu Eşitlik (2.13) ve (2.14)'ten,

$$L(Z, p, q) = \frac{\exp[-pn(Z) - q(n(G) - n(Z))][pn(Z) + q(n(G) - n(Z))]^{c(G)}}{c(G)!}$$

$$\prod_{x_i \in Z} \frac{pn(x_i)}{pn(Z) + q(n(G) - n(Z))} \prod_{x_i \notin Z} \frac{qn(x_i)}{pn(Z) + q(n(G) - n(Z))}$$

$$L(Z, p, q) = \frac{\exp[-pn(Z)-q(n(G)-n(Z))]}{c(G)!} p^{c(Z)} q^{c(G)-c(Z)} \prod_{x_i} n(x_i) \quad (2.15)$$

olarak elde edilir. Verilen bir Z bölgesi için p ve q 'nin en çok olabilirlik tahminleri,

$$\hat{p} = \frac{c(Z)}{n(Z)} \quad \text{ve} \quad \hat{q} = \frac{c(G)-c(Z)}{n(G)-n(Z)} \quad \text{biçimindedir. Bu tahminler Eşitlik (2.15)'te yerine}$$

yazılırsa Z bölgesi için olabilirlik fonksiyonu,

$$L(Z) = \begin{cases} \frac{\exp[-c(G)]}{c(G)!} \left(\frac{c(Z)}{n(Z)}\right)^{c(Z)} \left(\frac{c(G)-c(Z)}{n(G)-n(Z)}\right)^{c(G)-c(Z)} \prod_{x_i} n(x_i), & \frac{c(Z)}{n(Z)} > \frac{c(G)-c(Z)}{n(G)-n(Z)} \\ \frac{\exp[-c(G)]}{c(G)!} \left(\frac{c(G)}{n(G)}\right)^{c(G)} \prod_{x_i} n(x_i), & \text{ö. d.} \end{cases}$$

olarak bulunur. Yokluk hipotezi altındaki olabilirlik fonksiyonu ise,

$$L_0 \stackrel{\text{def}}{=} \sup_{p=q} \frac{e^{-pn(G)} p c(G)}{c(G)!} \prod_{x_i} n(x_i) = \frac{e^{-c(G)}}{c(G)!} \left(\frac{c(G)}{n(G)}\right)^{c(G)} \prod_{x_i} n(x_i).$$

biçimindedir. Böylece olabilirlik oranı λ aşağıdaki biçimde yazılabilir:

$$\lambda = \frac{\max_Z L(Z)}{L_0}.$$

Olabilirlik oranı, olası tüm alt bölgeler üzerinden maksimum yapılmaya çalışılır. Olabilirlik oranını maksimum yapan Z bölgesi en olası kümeyi gösterir (Kulldorff, 1997).

Poisson modeli için tarama istatistiğinin kullanıldığı bir örnek aşağıda verilmiştir:

Meksika'da 1973-1991 yılları arasında aşırı derecede meydana geldiği düşünülen beyin kanseri olaylarına göre bu bölgede iki boyutlu ve Poisson modeline uygun bir tarama istatistiği uygulanmış ve çeşitli sonuçlar elde edilmiştir.

Meksika'daki 32 şehir, SatScan programında değişen çaplı dairesel bir tarama penceresi ile taranmıştır. Dairenin risk altındaki popülasyonun en fazla %50'sini kapsayacak şekilde olması koşulu getirilmiştir. Beyin kanseri üzerinde etkili olduğu düşünülen yaş ve cinsiyet de analize dahil edilmiştir.

Analiz sonucunda anlamlı tek bir küme elde edilmiştir ($p = 0,030 < \alpha = 0,05$). Bu anlamlı kümede meydana gelen beyin kanseri sayısı $c(Z)$ 642 olup, beklenen değeri 583,2 olarak hesaplanmıştır. Bu kümede meydana gelen kanser sayısının Meksika'nın genelinde meydana gelen sayıdan %10 daha fazla olduğu görülmüştür (Kulldorff, 1999).

2.2.3 Normal Dağılım Modeli

Tarama istatistikleri genellikle, hastalıkların görülme sıklıkları ve ölüm sayısı gibi sayılabilen veriler için kullanılmaktadır. Bazı durumlarda ise, çocuklardaki kurşun düzeyi ve düşük doğum ağırlıkları gibi sürekli veriler ile ilgili kümelenmeler araştırılmak istenmektedir. Sürekli verilerin olasılık dağılımı için en çok normal dağılım kullanıldığından, tarama istatistikleri bu dağılım altında tanımlanmıştır.

Sürekli verilerin iki boyutlu analizleri ile ilgili ilk çalışma Kulldorff et al. (2009) tarafından yapılmıştır.

Normal dağılım modeli için gerekli tanımlamalar aşağıda verilmiştir:

x_i : olayların gözlem değeri $i = 1, \dots, N$

s : her gözlemin mekansal konumu $s = 1, \dots, S$

Her konum bir ya da daha fazla gözlem içereceğinden $S \leq N$ olur.

Her s konumu için, gözlem değerlerinin toplamı,

$$x_s = \sum_{i \in S} x_i$$

ile bulunurken her konumdaki gözlem sayısı n_s ile gösterilir.

Gözlenen değerler toplamı,

$$X = \sum_i x_i$$

her Z bölgesindeki gözlem sayısı,

$$n_z = \sum_{s \in Z} n_s$$

ile hesaplanırken bu gözlenen değerlerin toplamı,

$$x_z = \sum_{s \in Z} x_s$$

ile bulunur. Normal dağılım modeli için kurulan hipotezler aşağıdaki gibidir:

H_0 : Tüm gözlemler aynı normal dağılımdan gelmektedir.

H_1 : Tarama penceresi dışındaki gözlenen değerlerin ortalamasının pencerenin ortalamasından büyük ya da küçük olduğu bir tarama penceresi vardır.

Bu hipotezlerin testi için, yokluk hipotezi altında, $\hat{\mu} = \frac{X}{N}$ ve $\hat{\sigma}^2 = \frac{\sum_i (x_i - \hat{\mu})^2}{N}$ olmak üzere olabilirlik fonksiyonu,

$$L_0 = \prod_i \frac{1}{\hat{\sigma} \sqrt{2\pi}} e^{-\frac{(x_i - \hat{\mu})^2}{2\hat{\sigma}^2}}$$

ve log olabilirlik fonksiyonu,

$$\ln L_0 = -N \ln \left(\sqrt{2\pi} - N \ln(\hat{\sigma}) - \sum_i \frac{(x_i - \hat{\mu})^2}{2\hat{\sigma}^2} \right)$$

biçimindedir. Benzer şekilde seçenek hipotez altında olabilirlik fonksiyonu her Z bölgesi için ayrı ayrı hesaplanmaktadır. $\hat{\mu}_z = \frac{x_z}{n_z}$, pencere içindeki ortalama ve $\hat{\lambda}_z = \frac{(X-x_z)}{(N-n_z)}$ pencere dışındaki ortalama olmak üzere ortak varyans aşağıdaki gibi bulunur:

$$\hat{\sigma}^2 = \frac{1}{N} \left(\sum_{i \in Z} x_i^2 - 2x_z \hat{\mu}_z + n_z \hat{\mu}_z^2 + \sum_{i \notin Z} x_i^2 - 2(X - x_z) \hat{\lambda}_z + (N - n_z) \hat{\lambda}_z^2 \right).$$

Bu durumda Z bölgesi için log olabilirlik fonksiyonu,

$$\begin{aligned} \ln L(Z) &= -N \ln(\sqrt{2\pi}) - N \ln \left(\sqrt{\hat{\sigma}_z^2} \right) \\ &\quad - \frac{1}{2\hat{\sigma}_z^2} \left(\sum_{i \in Z} x_i^2 - 2x_z \hat{\mu}_z + n_z \hat{\mu}_z^2 + \sum_{i \notin Z} x_i^2 - 2(X - x_z) \hat{\lambda}_z + (N - n_z) \hat{\lambda}_z^2 \right) \end{aligned}$$

şeklinde hesaplanır. Bu hesaplama aşağıdaki gibi sadeleştirilebilir:

$$\ln L(z) = -N \ln(\sqrt{2\pi}) - N \ln \left(\sqrt{\sigma_z^2} \right) - \frac{N}{2}.$$

Hipotezlerin testi için log olabilirlik oranı,

$$\begin{aligned} \lambda &= \frac{\max_z L(Z)}{L_0} \\ \ln \lambda &= \max_z (\ln L(Z) - \ln L_0) \\ &= \max_z \left(-N \ln(\sqrt{2\pi}) - N \ln \left(\sqrt{\hat{\sigma}_z^2} \right) - \frac{N}{2} + N \ln(\sqrt{2\pi}) + N \ln(\hat{\sigma}) + \sum_i \frac{(x_i - \hat{\mu})^2}{2\hat{\sigma}^2} \right) \\ &= \max_z \left(N \ln(\hat{\sigma}) + \sum_i \frac{(x_i - \hat{\mu})^2}{2\hat{\sigma}^2} - \frac{N}{2} - N \ln \left(\sqrt{\hat{\sigma}_z^2} \right) \right) \end{aligned} \quad (2.16)$$

olarak bulunmaktadır. Eşitlik (2.16)'daki ifade Z 'ye bağlı olduğundan en olası kümenin seçenek hipotez altında varyansı minimum yapan küme olduğu söylenebilir. Tarama istatistikleri çok sayıda kesişen tarama penceresi üzerinden tanımlandığından, her Z bölgesi (tarama penceresi) için olabilirlik oranı hesaplanır ve test istatistiği değeri tüm tarama pencereleri üzerinden hesaplanan olabilirlik oranının maksimumu olarak belirlenir (Kulldorff et al., 2009).

Normal dağılım modeli için bir örnek aşağıda verilmiştir:

New York Sağlık Departmanı tarafından 2004 yılında toplanan veriler üzerinde normal dağılımlı tarama istatistiği ile dairesel ve değişen boyutlu tarama penceresi uygulanmış ve bazı sonuçlar elde edilmiştir. Amaç, bebek ölüm oranını etkilediği düşünülen düşük doğum ağırlığının iki boyutlu olarak araştırılması ve anlamlılığının test edilmesidir.

Burada:

Düşük doğum ağırlığı: <2500 gr

Çok düşük doğum ağırlığı: <1500 gr

olarak tanımlanmıştır.

Doğumlar, annelerin bulunduğu yerlerin posta koduna göre coğrafi olarak kodlanmıştır.

Değişken tanımlamaları ise şu şekildedir:

x_i : doğum ağırlıkları ($i = 1, \dots, N$)

s : her gözlem için mekansal konum ($s = 1, \dots, S$)

Mekansal enlem koordinatı: $lat(s)$

Mekansal boylam koordinatı : $long(s)$

Her mekansal konum bir ya da daha çok gözlem içerir ($S \leq N$).

$x_s = \sum_{i \in S} x_i$: her s konumundaki gözlenen değerler toplamı

n_s : her konumdaki gözlem sayısı

$X = \sum_i x_i$: tüm gözlenen değerlerin toplamı

$n_z = \sum_{s \in Z} n_s$: Z bölgesindeki gözlem sayısı

$x_z = \sum_{s \in Z} x_s$: Z bölgesindeki gözlenen değerlerin toplamı

Hesaplanan olabilirlik oran istatistiklerine göre, 2 tane önemli küme tespit edilmiştir: İlk küme, 125,8 olabilirlik oranı ile 61 tane posta kodu içermektedir. Bu küme doğu Brooklyn ve güney Queens'i kapsamaktadır. Bu kümedeki doğum ağırlığı şehrin geri kalanından ortalama 60 gr daha azdır ($p < 0,001$). İkinci küme ise 62,7 olabilirlik oranı ile 29 posta kodundan oluşmaktadır ($p < 0,001$). Kuzey Manhattan ve Güney Bronx'u kapsamaktadır. Bu kümede de doğum ağırlıkları şehrin geri kalanından ortalama 52 gr daha azdır (Kulldorff et al., 2009).

2.2.4 Üstel Dağılım Modeli

İki boyutlu tarama istatistiklerinde üstel dağılım modeli genellikle sağ kalım ve hastalıkların ilerlemesi gibi sağlık olaylarının coğrafi dağılımını analiz etmek için kullanılır. Örneğin Karjalainen (1990) Finlandiya'daki kanser hastalarının sağ kalım sürelerinin coğrafi dağılımı üzerine çalışmıştır. Jack et al. (2003) İngiltere'deki akciğer hastalarının sağ kalım sürelerinin coğrafi farklılıkları üzerine araştırma yapmışlardır.

Üstel dağılıma dayalı iki boyutlu tarama istatistiklerinde, Z bölgesindeki her bir bireyin sağ kalım süresi θ_z ortalaması ile üstel dağılıma sahipken, bu bölgenin dışında kalan bireylerin sağ kalım süreleri ise θ ortalaması ile üstel dağılıma sahiptir. Bu model altında tanımlanan hipotezler aşağıdaki gibidir:

$H_0: \theta_z = \theta$ (sağ kalım süresinin farklı olduğu bir bölge yoktur)

$H_1: \theta_z < \theta$ (sağ kalım süresi bölge dışından daha küçük olan en az bir bölge vardır)

ya da

$H_1: \theta_z > \theta$ (sağ kalım süresi bölge dışından daha büyük olan en az bir bölge vardır)

ya da

$H_1: \theta_Z \neq \theta$ (sağ kalım süresinin bölge dışından daha büyük ya da küçük olduğu en az bir bölge vardır)

Seçenek hipotez altında θ_Z ve θ ortalamaları farklı bölgeler için farklı değerler alırlar.

N , çalışmadaki birey sayısı olmak üzere, i . bireyin yaşam süresi T_i , sansürlenmiş sabit zamanı ise L_i ile gösterilsin.

Üstel dağılım gösteren T_i yaşam sürelerinin olasılık fonksiyonu aşağıdaki gibidir:

$$f(t_i) = \frac{1}{\theta} e^{-\frac{t_i}{\theta}}, \quad t_i > 0$$
$$= 0, \quad t_i \leq 0.$$

Yaşam süresi T_i , yalnızca $T_i \leq L_i$ olduğu durumlarda gözlenir. Diğer durumda ise, sağ kalım süresi sansürlenmiş olur. Bu durumda gözlem yapılan zaman,

$$y_i = \min(T_i, L_i)$$

ile ifade edilebilir. δ_i yaşam süresinin sansürlü olup olmadığını göstermek üzere, $T_i \leq L_i$ olduğu durumda $\delta_i = 1$ iken diğer durumda ise 0'a eşit olur. Benzer şekilde gerekli tanımlamalar aşağıda verilmiştir:

$r_Z = \sum_{i \in Z} \delta_i$: Z bölgesinde sansürlenmemiş birey sayısı,

$r = \sum_{i \notin Z} \delta_i$: Z bölgesi dışında sansürlenmemiş birey sayısı,

n_Z : Z bölgesindeki birey sayısı,

n : Z bölgesi dışındaki birey sayısı,

$N = n_Z + n$: G çalışma alanındaki toplam birey sayısı

$R = r_Z + r$: G çalışma alanındaki sansürlenmemiş toplam birey sayısı.

Keyfi bir Z bölgesi için, $\theta_Z \neq \theta$ seçenek hipotezi altında olabilirlik fonksiyonu aşağıda verilmiştir:

$$L(Z, \theta_Z, \theta) = \prod_{i \in Z} \frac{1}{(\theta_Z)^{\delta_i}} e^{-\frac{t_i \delta_i}{\theta_Z}} e^{-\frac{l_i(1-\delta_i)}{\theta_Z}} \times \prod_{i \notin Z} \frac{1}{(\theta)^{\delta_i}} e^{-\frac{t_i \delta_i}{\theta}} e^{-\frac{l_i(1-\delta_i)}{\theta}}$$

$$= \frac{1}{(\theta_Z)^{r_Z}} e^{-\sum_{i \in Z} \frac{y_i}{\theta_Z}} \frac{1}{(\theta)^r} e^{-\sum_{i \notin Z} \frac{y_i}{\theta}}.$$

Keyfi bir Z bölgesindeki θ_Z ve θ parametrelerinin en çok olabilirlik tahmin edicileri $\hat{\theta}_Z = \frac{r_Z}{\sum_{i \in Z} y_i}$ ve $\hat{\theta} = \frac{r}{\sum_{i \notin Z} y_i}$ olurken yokluk hipotezine karşı seçenek hipotezi için olabilirlik oranı,

$$\lambda = \frac{\max_Z L(Z)}{L_0} \quad (2.17)$$

İken, yokluk hipotezi altındaki olabilirlik fonksiyonu L_0 ,

$$L_0 = \frac{1}{(\hat{\theta}_G)^R} e^{-\sum_{i \in G} \frac{y_i}{\hat{\theta}_G}} = \left(\frac{R}{\sum_{i \in G} y_i} \right)^R e^{-R}$$

biçiminde bulunur. Eşitlik (2.17)'de verilen olabilirlik oranını maksimum yapan Z bölgesi en olası kümeyi gösterir.

Sansürlü birey olduğu durumda seçenek hipotez altında olabilirlik fonksiyonu,

$$L'(Z) = \max_Z \frac{1}{(\hat{\theta}_Z)^{r_Z}} e^{-\sum_{i \in Z} \frac{y_i}{\hat{\theta}_Z}} \frac{1}{(\theta)^{r_{out}}} e^{-\sum_{i \notin Z} \frac{y_i}{\theta}}$$

$$= \max_Z \left(\frac{r_Z}{\sum_{i \in Z} y_i} \right)^{r_Z} e^{-r_Z} \left(\frac{r}{\sum_{i \notin Z} y_i} \right)^r e^{-r} \quad (2.18)$$

olarak bulunur.

$\theta_Z < \theta$ seçenek hipotezi altında Eşitlik (2.18)'de verilen fonksiyon $I\left(\frac{r_Z}{\sum_{i \in Z} y_i} < \frac{r}{\sum_{i \notin Z} y_i}\right)$ ile çarpılırken, $\theta_Z > \theta$ seçenek hipotezi altında Eşitlik (2.18)'de verilen fonksiyon $I\left(\frac{r_Z}{\sum_{i \in Z} y_i} > \frac{r}{\sum_{i \notin Z} y_i}\right)$ ile çarpılır.

Sansürlü birey olması durumunda yokluk hipotezi altındaki olabilirlik fonksiyonu Eşitlik (2.17)'deki gibidir. Bu durumda olabilirlik oranı,

$$\lambda = \frac{\max_Z L'(Z)}{L_0}$$

biçimindedir. Sansürlü birey olmadığı durumda ise,

$$L''(Z) = \max_Z \left(\frac{n_Z}{\sum_{i \in Z} y_i} \right)^n e^{-n_Z} \left(\frac{n}{\sum_{i \notin Z} y_i} \right)^n e^{-n}$$

$$L_0 = \left(\frac{N}{\sum_{i \in G} y_i} \right)^N e^{-N}$$

ile hesaplanır. Bu durumda olabilirlik oranı aşağıdaki gibidir:

$$\lambda = \frac{\max_Z L''(Z)}{L_0}$$

Benzer şekilde olabilirlik oranı, olası tüm alt bölgeler üzerinden maksimum yapılmaya çalışılır. Olabilirlik oranını maksimum yapan Z bölgesi en olası kümeyi gösterir.

Üstel dağılım modeli altında tarama istatistiğinin kullanıldığı bir örnek aşağıda verilmiştir:

Amerika'nın Connecticut eyaletinde 1984-1995 yılları arasında risk altındaki 20 yaş üstü erkeklerden prostat kanseri olan 20598 bireyin kayıtları üzerinden iki boyutlu tarama istatistiklerinde üstel dağılım modeli uygulanmıştır. Bu bireylerin 1325'inin yaşam süresi bilgisi eksik olduğundan ve 212 bireyin de teşhis konduğu gün ölmesinden dolayı toplam kullanılabilir veri sayısı 19061'e düşmüştür. Bu bireylerin de 10308'inin ölüm zamanları kayıtlı olurken geri kalan 8753 birey ise sansürlenmiştir. Eldeki verilere göre yapılan analiz sonucunda, en olası Batı Connecticut bölgesinde olup, kümede bulunan bireylerin diğer yerlerdeki bireylere göre ölüm riski %29 daha fazla olarak bulunmuştur ($p = 0,001$) (Huang et al., 2007).

2.2.5 Katlı Terimli (Multinomial) Dağılım Modeli

Coğrafik kümelenmelerin tespitinde iki boyutlu tarama istatistiği Bernoulli, Poisson, normal, üstel gibi birçok model için geliştirilmiştir. Bu modellerden bir diğeri de katlı terimli dağılım modelidir. Örneğin, bir hastalığın farklı türlerinin olduğu bir bölgede bir araştırmacı hastalık türünün dağılımının istatistiksel olarak anlamlı bir şekilde

bölgenin geri kalan kısmından farklı olduğu bir kümenin bulunması için bir çalışma yapılabilir.

I tane alt kümeden oluşan (il, posta kodu gibi) bir bölgedeki bir hastalığın K tane farklı türü olduğunda c_{ik} , i . alt kümedeki k . tür hastalığın gözlem sayısı olarak tanımlanmaktadır ($k = 1, \dots, K$, $i = 1, \dots, I$).

p_k , k türünün Z bölgesi içinde olması olasılığını gösterirken q_k ise bu türün Z bölgesi dışında olması olasılığını göstermektedir. Aynı zamanda bu olasılıkların k üzerinden toplamları 1'e eşit olmaktadır ($\sum_k p_k = \sum_k q_k = 1$). Her tür için olasılıkların eşit olduğu varsayımı ile kurulan yokluk hipotezi,

$$H_0: p_1 = q_1, \dots, p_k = q_k$$

$$H_1: \text{en az bir } p_k \neq q_k, \quad k = 1, 2, \dots, K$$

şeklinde iken seçenek hipotez ise, en az bir türün olasılığının diğerlerinden farklı olduğu şeklinde kurulmaktadır.

$c_i = \sum_k c_{ik}$: alt bölgelerdeki toplam gözlenen olay sayısı,

$C_k = \sum_i c_{ik}$: k türünde görülen toplam gözlem sayısı

$C = \sum_k \sum_i c_{ik}$: tüm çalışma alanındaki toplam gözlenen olay sayısı.

Yokluk hipotezi altında olabilirlik fonksiyonu,

$$L_0 = \prod_k \prod_i \hat{p}_{o_k}^{c_{ik}} = \prod_k \left(\frac{C_k}{C} \right)^{\sum_i c_{ik}} = \prod_k \left(\frac{C_k}{C} \right)^{C_k}$$

ile hesaplanırken burada, $\hat{p}_{o_k} = \frac{C_k}{C} = \hat{q}_{o_k}$, yokluk hipotezi altında $p_k = q_k$ 'nin en çok olabilirlik tahminidir.

Seçenek hipotez altında olabilirlik fonksiyonu ise,

$$L(Z) = \prod_k \left(\prod_{i \in Z} \hat{p}_k^{c_{ik}} \prod_{i \notin Z} \hat{q}_k^{c_{ik}} \right)$$

ile hesaplanırken benzer şekilde burada da, p_k ve q_k 'nin en çok olabilirlik tahminleri sırasıyla $\hat{p}_k = C_k(Z)/C(Z)$, ve $\hat{q}_k = (C_k - C_k(Z))/(C - C(Z))$ 'dir.

Olabilirlik oranı,

$$\lambda = \frac{\max_z L(Z)}{L_0}$$

biçimindedir. Olabilirlik oranı, olası tüm alt bölgeler üzerinden maksimum yapılmaya çalışılır. Olabilirlik oranını maksimum yapan Z bölgesi en olası kümeyi gösterir.

Bu bölümde çeşitli modeller için tanımlanan iki boyutlu tarama istatistikleri, tarama bölgesi Z 'nin aralık olarak alınması durumunda tek boyutlu tarama istatistiklerine dönüşür.

Katlı terimli dağılım modeli kullanımı ile ilgili aşağıdaki örnek verilebilir:

İngiltere'de 1999-2001 yılları arasında iki şehirdeki 18 yaş altı menenjit hastalığı olan 475 öğrenci seçilmiş ve bu öğrenciler menenjitin 5 farklı türüne göre (ST-213 complex, ST-22 complex, ST-23 complex / cluster A3, ST -41 / 44 complex / lineage 3, ST-53 complex) gruplandırılmıştır. Bu çalışmada amaç, iki şehirde menenjit türleri dağılımında rasgele olmayan mekansal kümelenmelerin belirlenmesidir (Jung et al., 2010).

Analiz sonucunda elde edilen sonuçlar aşağıdaki çizelgede verilmiştir:

	Yarıçap (m)	Gözlem Sayısı	Görel Risk (Her bir tür için)	Olabilirlik Oranı	p-değeri
1. Küme	4948	56	(0,21; 0,82; 0,00; 1,07; 3,00)	18,10	0,001
2. Küme	1745	18	(4,94; 0,23; 0,55; 0,62; 0,00)	14,44	0,033
3. Küme	9546	37	(0,51; 1,83; 0,00; 1,51; 0,00)	14,10	0,041

Çizelge 2.1. Katlı Terimli Dağılım Modeli Örneği için Bulunan Anlamlı Kümeler

2.2.2 İki Boyutlu Tarama Penceresinin Şekli

İki boyutlu tarama istatistiğindeki en önemli unsurlardan biri tarama penceresinin şekli ve boyutudur. Naus (1965b) sabit şekil ve büyüklükteki dikdörtgen bir tarama penceresi kullanırken, Loader (1991) değişen büyüklüklü bir dikdörtgen tarama penceresi kullanmıştır. Alm (1997,1998) daire, elips ve üçgen tarama pencereleri üzerine çalışmalar yapmıştır. Kulldorff (1997) kendi çalışmasında değişen çaplı daire kullanmıştır.

Tarama penceresinin şekli uygulamaya bağlı olmasına rağmen genellikle bir gözlem üzerinde merkezlenen ve yarıçapı 0'dan belirlenen bir üst sınıra kadar değişen bir daire kullanılmaktadır. Üst sınır çoğu kez tüm gözlemlerin en fazla %50'sinin içerildiği bir daire genişliği olarak belirlenir. Bunun yanında üst sınır 10 kilometre gibi bir uzaklık olarak da belirlenebilir. Tek gözlemler anlamlı kabul edilmez, yorumlanmazlar (Kulldorff, 1997).

2.3 Üç Boyutlu (Uzay-Zaman) Tarama İstatistiği

Üç boyutlu tarama istatistiği, iki boyutlu tarama istatistiğinin genelleştirilmiş biçimidir. Üç boyutlu tarama istatistiğinde tarama penceresi dairesel bir taban üzerinde yüksekliği zamanı gösterecek şekilde silindirik bir pencere olarak tanımlansın. Silindirin tabanı, değişen yarıçaplı bir daire olup merkez noktası ise çalışma alanındaki mümkün tüm konumlar üzerindedir. Silindirin yüksekliği ise, tüm çalışma süresinin yarısının ya da daha az kısmının tüm olası zaman aralığını yansıtır. Daha sonra tanımlanan silindirik pencere tüm olası uzay üzerinde tüm olası zaman aralıklarını da tarayarak hareket eder.

Üç boyutlu tarama istatistiğinde kurulan hipotezler aşağıdaki gibidir:

H_0 : üç boyutta kümelenme yoktur.

H_1 : üç boyutta kümelenme vardır.

Kulldorff et al. 2005 yılındaki üç boyutlu tarama istatistiği ile ilgili çalışması Kulldorff et al. 2001 yılındaki çalışmasına dayanmaktadır. Bu çalışmaya göre üç boyutlu tarama istatistiği için kullanılan tanımlamalar aşağıda verilmiştir:

d günü içinde z bölgesinde gözlenen olay sayısı c_{zd} olmak üzere toplam gözlenen olay sayısı,

$$C = \sum_z \sum_d c_{zd}$$

iken her bölge ve gün için beklenen olay sayısı,

$$\mu_{zd} = \frac{1}{C} \left(\sum_z c_{zd} \right) \left(\sum_d c_{zd} \right)$$

formülü ile hesaplanır. A silindirik tarama penceresi içerisindeki beklenen olay sayısı ise,

$$\mu_A = \sum_{(z,d) \in A} \mu_{zd}$$

eşitliği ile hesaplanır.

C_A , silindir içinde belirli bir özelliği gösteren olayların sayısı olmak üzere, boyutlar bakımından bir etkileşim olmadığında C_A , μ_A ortalamalı hipergeometrik dağılım göstermektedir. Bu dağılımın olasılık fonksiyonu aşağıdaki gibidir:

$$P(C_A) = \frac{\binom{\sum_{z \in A} c_{zd}}{C_A} \binom{C - \sum_{z \in A} c_{zd}}{\sum_{d \in A} c_{zd} - C_A}}{\binom{C}{\sum_{d \in A} c_{zd}}}$$

A tarama penceresi içindeki belirli bir özelliği gösteren olaylar için hipergeometrik dağılımın Poisson dağılımına yaklaşımından yararlanarak genelleştirilmiş Poisson olabilirlik oranı istatistiği aşağıdaki gibi hesaplanır:

$$GLR = \left(\frac{C_A}{\mu_A} \right)^{C_A} \left(\frac{C - C_A}{C - \mu_A} \right)^{C - C_A}$$

Yukarıda verilen olabilirlik oran istatistiğini maksimum olan silindir en olası kümeyi göstermektedir.

Üç boyutlu tarama istatistiğinin kullanımı ile ilgili aşağıdaki örnek verilebilir:

New York'taki acil servislere gelen hastalar üzerinde üç boyutlu tarama istatistiği kullanılarak uzay-zaman boyutunda hastalık kümelenmeleri araştırılmıştır. Çalışmanın amacı ilgilenilen hastalıkla ilgili acil servislere gelenlerde beklenmedik artışların olup olmadığının araştırılmasıdır. Zaman aralığı 15 Kasım 2001 ile 14 Kasım 2002 yıllarını kapsamaktadır. Bu yıllar arasında acil servislere bağırsak sorunu nedeniyle gelen hastaların yaşadıkları yerin koordinatları ve hastanelerin koordinatları kullanılarak iki farklı analiz yapılmıştır. 38 hastaneye karşılık 183 yerleşim yeri analizde yer almaktadır.

Hastane koordinatlarına göre silindirik tarama penceresi kullanılarak yapılan analiz sonucunda, en olası kümede bağırsak sorunu nedeniyle acil servislere başvuranların sayısının 21 Kasım 2001 tarihinde artış olduğu bulunmuştur. Gözlenen hasta sayısı 101 iken, beklenen hasta sayısı ise 73,6 olarak hesaplanmıştır (Kulldorff et al., 2005).

İzleyen bölümde, tek, iki, üç boyutlu tarama istatistiklerinde kurulan hipotezlerin testi ve testin gücü ile ilgili genel bilgi verilmiştir.

3. HİPOTEZ TESTİ ve TESTİN GÜCÜ

Tarama istatistiğinin üç temel özelliği, çalışma alanının geometrik şekli, yokluk hipotezi altında türetilen ve belirli bir özelliği gösteren olayların olasılık dağılımı ve tarama penceresinin şekli ve büyüklüğü olarak tanımlanmaktadır.

Tarama penceresi genellikle sabit büyüklük ve boyutlardaki bir aralık, alan ya da hacim olarak belirlenebilir. Bu pencere, tarama alanında hareket ettikçe $Z \subset G$ bölgelerinin tamamından oluşan Z yığınlarını tanımlamaktadır. Tarama penceresi büyüklüğü sabit olduğunda tarama istatistiği, tüm tarama pencereleri içindeki olayların maksimum sayısına eşittir (Kulldorff, 1999).

$$S'_Z = \max_{Z \in \mathcal{Z}} C(Z)$$

Eğer değişen büyüklük ve boyuttaki bir tarama penceresi kullanılırsa bu durumda tarama istatistiği,

$$S_Z = \max_{Z \in \mathcal{Z}} \frac{L(Z)}{L_0}$$

biçiminde tanımlanabilir. Burada $L(Z)$, Z bölgesinin olasılık fonksiyonu olup L_0 ise yokluk hipotezi altında hesaplanan olasılık fonksiyonudur.

Örneğin Bernoulli modelinde,

$$C(G \setminus Z) = C(G) - C(Z)$$

ve

$$n(G \setminus Z) = n(G) - n(Z)$$

olmak üzere tarama istatistiği,

$\frac{C(Z)}{n(Z)} > \frac{C(G \setminus Z)}{n(G \setminus Z)}$ olduğu durumda aşağıdaki gibi bulunur:

$$\frac{L(Z)}{L_0} = \frac{\left(\frac{C(Z)}{n(Z)}\right)^{C(Z)} \left(1 - \frac{C(Z)}{n(Z)}\right)^{n(Z)-C(Z)} \left(\frac{C(G \setminus Z)}{n(G \setminus Z)}\right)^{C(G \setminus Z)} \left(1 - \frac{C(G \setminus Z)}{n(G \setminus Z)}\right)^{n(G \setminus Z)-C(G \setminus Z)}}{\left(\frac{C(G)}{n(G)}\right)^{C(G)} \left(1 - \frac{C(G)}{n(G)}\right)^{n(G)-C(G)}}$$

Diğer durumlarda ise $L(Z) = 1$ olur.

Poisson modelinde ise,

$\frac{c(Z)}{n(Z)} > \frac{c(G \setminus Z)}{n(G \setminus Z)}$ olduğu durumda test istatistiği aşağıdaki gibidir:

$$\frac{L(Z)}{L_0} = \frac{\left(\frac{C(Z)}{n(Z)}\right)^{c(Z)} \left(\frac{C(G \setminus Z)}{n(G \setminus Z)}\right)^{c(G \setminus Z)}}{\left(\frac{C(G)}{n(G)}\right)^{c(G)}}.$$

Diğer durumlarda ise $L(Z) = 1$ olur.

$\frac{c(Z)}{n(Z)} > \frac{c(G \setminus Z)}{n(G \setminus Z)}$ koşulunun sağlanması tarama penceresi içindeki olayların beklenen değerinin pencere dışındaki olayların beklenen değerinden daha büyük olduğunu göstermektedir.

Testin gücü ile ilgili Wallenstein et al. (1993), Sahu et al. (1993) vb. çalışmalar literatürde yer almaktadır. Bu çalışmalarda test istatistiğinin dağılımı ve testin gücü ile ilgili bazı yaklaşımlar verilmektedir.

Test istatistiğinin dağılımı analitik olarak bulunamadığından test istatistiğinin değeri ve yokluk hipotezinin anlamlılığının bulunmasında Monte Carlo benzetim tekniği kullanılmaktadır. Bu nedenle diğer bölümde Monte Carlo benzetiminden bahsedilecektir.

4. MONTE CARLO BENTEZİMİ ve ALGORİTMALAR

Monte-Carlo benzetimi, ilk kez II. Dünya savaşı sonrasında orduyu ve savaşı yönetmek maksadıyla kullanılmış, daha sonra bu çalışmalar modern benzetimin temelini oluşturmuştur. Benzetim 1940'dan günümüze; olasılıksal finansal planlama, sigorta değerlendirmesi, yatırım modelleri ve talep tahminleri gibi birçok alana başarıyla uygulanmıştır.

Monte Carlo metodu tekbiçimli dağılımdan rasgele değişkenler elde etmek ve bunları uygun bir şekilde ilgilenilen dağılıma taşımaktır. Bir tekbiçimli dağılım, değişken değerlerinin özel bir alanla sınırlandırılmış olması ve eşit şanslara sahip olması durumunda veya aynı olasılıklara sahip olması durumunda mevcuttur. Çoğunlukla bu tekbiçimli rasgele değişkenlerden rasgele sayılar olarak bahsedilir. Bu sayılar iki önemli şartı sağlamaktadırlar:

- 1- Bütün değerlerin meydana gelmesi eşit şansa sahiptir.
- 2- Yeni değerler bütün önceki elde edilen değerlerden bağımsızdır.

Diğer bir deyişle, eğer biz tek haneli rasgele sayıları elde ediyorsak, sıfırdan dokuzaya kadar olan sayıların elde edilme olasılığı geçmişte elde edilen sayılardan bağımsız olarak birbirine eşittir. Yani 3 sayısının %10'luk elde edilme olasılığı 0-9 arasındaki sayıların hepsiyle aynıdır. Benzetim modeli için rasgele tekbiçimli sayıların ihtiyacımız olan rastgele değişkenlere çevrilmesini çeşitli yollarla meydana getirebiliriz. En basiti sayıları alt kümeler şeklinde özel olaylara tahsis etmektir. Öyle ki alt kümedeki sayıların olasılıklarının birleşimi modellenen olayın olasılığına tam olarak karşı gelmelidir (Öztürk, 2004).

Benzetim modelleri matematiksel bir formülasyon içermediğinden dolayı model çözümlerinde rasgele sayılardan yararlanır. Rasgele değişkenler için hazırlanan olasılık dağılımları rasgele sayılar tablosuna göre düzenlenir. Bunun yapılabilmesi için olasılık dağılımlarından birikimli (kümülatifli) olasılık dağılımları elde edilir. Bunlar elde edildikten sonra oluşturulan Monte-Carlo sayı aralıklarına göre belirli bir düzen içerisinde, göre her aralığa karşılık gelecek rasgele değişkenin değeri belirlenir.

Rasgele sayılar tablosundan oluşturulan Monte-Carlo sayılarına göre (birikimli olasılıklara bağlı olarak oluşturulan Monte-Carlo sayıları) seçilecek sayılar belirli bir

düzene göre alınmalıdır. Örneğin, ikinci satır ikinci sütunun ilk iki rakamı veya son iki rakamı sağa doğru veya aşağıya doğru gözlem sayısına dayanarak seçilir. Aylar itibariyle bir değerlendirme yapıldığında 12 adet rasgele sayı seçilecek bu sayılara göre Monte-Carlo sayılarının düştüğü aralık belirlenip karşılık gelen rasgele değişken bulunacaktır. Bundan da anlaşılacağı üzere rasgele sayılar tablosu benzetim modellerinin olasılıksal özellikli Monte-Carlo benzetim tekniği ile çözümünde çok önemli bir vazife görmektedir (Kahveci, 2002).

Tarama istatistiklerinde, test istatistiği değeri, seçenek hipotez altında tüm bölgeler üzerinden maksimum olabilirlik oranının hesaplanmasını gerektirir. Bölgelerin sayısı sonsuz olabileceğinden bu hesaplama oldukça zordur. Bu zorluğu giderebilmek için gözlem sayısı sonlu olduğundan hesaplanacak olabilirlik oranı sayısı da sonlu olur. Test istatistiği hesaplandığında, sonuca ulaşmak kolaylaşır. Test istatistiğinin dağılımını bulmak için Monte-Carlo benzetiminden yararlanılır. Bu teknik ilk olarak Turnbull(1990) tarafından tarama istatistikleri için kullanılmıştır. Elimizdeki verilerin ölçümleri bilindiğinden, toplam gözlem sayısına bağlı olarak yokluk hipotezi altında oluşturulan veri setinden yeni veri setleri oluşturulur. Örneğin, 9999 kez yapılan yineleme ile eğer gerçek veri setinden elde edilen test istatistiği değeri rasgele oluşturulan veri setlerinden elde edilen test istatistiklerinin en yüksek 500 tanesi içindeyse, o test istatistiği $\alpha = 0,05$ 'te anlamlı demektir.

En olası kümeye ek olarak, yüksek olabilirlik değerine sahip ikincil kümelerin bulunması da istenebilir. Bazı durumlarda ikincil kümelerin kapsadıkları olaylar en olası küme ile aynı bölgede bulunmakta ve kümeler çakışmaktadır. Bu durumda ikincil kümenin yorumlanmasına gerek yoktur. Eğer bu kümeler çalışma alanında farklı yerlerde bulunur ve çakışmazlarsa o zaman bu kümelerin yorumlamaları yapılabilir (Kulldorff, 1997).

Tarama istatistiklerinin uygulamasında test işlemi dört adımdan oluşmaktadır:

- 1) Gerçek veriler için test istatistiğinin değerinin hesaplanması.
- 2) Yokluk hipotezi altında çok sayıda rasgele veri setlerinin oluşturulması.
- 3) Her bir rasgele veri seti için test istatistiğinin değerinin bulunması.
- 4) Gerçek ve rasgele veri setlerindeki test istatistiklerinin değerlerinin sıralanması ve gerçek veri setinden oluşturulan test istatistiğinin değerinin sıra sayısının

not edilmesi. Eğer gerçek veri setindeki test istatistiğinin sıra sayısı en yüksek α yüzdesi içindeyse, yokluk hipotezi α anlamlılık düzeyinde reddedilir (Kulldorff, 1999).

Bunu izleyen kesimde tarama istatistiklerinde kullanılan çeşitli algoritmalara yer verilmiştir.

4.1. Algoritmalar

Algoritma 1: Dairesel Pencere. Sabit Büyüklük. Homojen Poisson Süreci

1. Belirli bir özelliği gösteren olay için (x_i, x_j) ve (y_i, y_j) konumları, aralarındaki uzaklık d uzunluğundan fazla olmayacak şekilde tanımlanır.
2. (x_i, x_j) ve (y_i, y_j) çember üzerinde bulunacak şekilde d çaplı iki daire çizilir.
3. Her iki dairenin üzerindeki ve içindeki olay sayısı bulunup, n değeri bu sayıların maksimumu olarak tanımlanır.
4. 1'den 3'e kadar her adım tüm ilişkili olaylar için tekrarlanıp, n değerleri içindeki en büyük değer tarama istatistiği olarak belirlenir.

$n(G)$, toplam olay sayısı ve $O(n(G))$, algoritmanın performansının analizi yapılan veri setinin büyüklüğüyle doğru orantılı olduğunu göstermesi iken 3. adıma bir kere gelinmesi karmaşıklığı $O(n(G))$ kadardır. 1. ve 3. adımlar R tane Monte Carlo rasgele sayı üretiminin her biri için $O(n(G)^2)$ kere tekrarlanır. Böylece toplam karmaşıklığı $O(n(G)^3)$ olur. Burada yer alan $O(\cdot)$ ifadesi algoritma karmaşıklığını gösteren bir fonksiyondur.

Algoritma 2: Dairesel Pencere. Sabit Büyüklük. Homojen Poisson Süreci

$n(G)$ çok büyük olduğunda daha etkili olan algoritmadır.

1. Belirli bir özelliği gösteren bir olayın x (x_i, x_j) konumu belirlenir ve x merkezli, d yarıçaplı büyük bir daire çizilir. x arasındaki açı 0° olacak şekilde daire üzerindeki keyfi bir yerde x_0 konumu belirlenir.
2. Büyük daire içine yarıçapı $\frac{d}{2}$ olan daha küçük bir daire çizilir. Küçük daire, büyük daire içinde saat yönünde hareket ettirilir. Bu hareket sırasında x konumu her zaman küçük çember üzerinde bulunmalı. Her iki çember

üzerinde yer alacak şekilde bir x_a noktası tanımlanır. Burada a , 0^0 ile 360^0 arasında olup x ile x_a arasındaki açıyı göstermektedir.

3. Büyük dairenin üzerindeki ve içindeki her olay için, küçük daire hareket ettikçe içine giren ya da çıkan olaylara göre x ile x_a arasındaki açılar kaydedilir. Bu açılar, daireye giriş ve çıkışlarla ilgili olarak küçükten büyüğe doğru sıralanır.
4. Çemberi üzerinde x ile x_0 bulunan küçük daire içindeki olaylar sayılır. Daha sonra küçükten büyüğe doğru sıralanan açılarda, daireye veri girişi olanlara 1 ekleyip, veri çıkışı ile ilgili olanlardan 1 çıkarılır. En büyük sayı n olarak tanımlanır.
5. Her olay için 1'den 4'e kadar tüm adımlar tekrarlanır. En büyük n değeri tarama istatistiği olarak tanımlanır.
6. Her Monte Carlo sayı üretimi için 1'den 5'e kadar tüm adımlar tekrarlanır.

3. adımda sıralama $O(n(G)\log n(G))$ kez yapılırken 2. ve 4. adımlar $O(n(G))$ kez tekrarlanır. 1'den 5'e kadar olan adımlarda ise her Monte Carlo sayı üretimi için $n(G)$ defa iterasyon yapılır. Bu durumda toplam karmaşıklık, $O(Rn(G))[O(n(G)) + O(n(G)\log n(G)) + O(n(G))] = O(Rn(G)^2\log n(G))$ olarak hesaplanır.

Algoritma 3: Dairesel Pencere. Sabit Büyüklük. Homojen Poisson Süreci

1. Belirli bir özelliği gösteren iki olay için (x_i, x_j) ve (y_i, y_j) konumları belirlenir. Bu iki konuma eşit uzaklıkta olan bir L doğrusu çizilir. Bu doğrunun uçlarından biri sol uç olarak tanımlanır.
2. Geriye kalan her z olayı için, çemberin üzerinde x, y, z olacak şekilde bir daire çizilir. L , ağırlık merkezi z olan daire üzerindedir ve ağırlık merkezi sola doğru kaydıkça z daireye girer ya da çıkar.
3. L üzerindeki daire merkezleri sağdan sola doğru sıralanır. Sıralama yapılırken her bir merkezin z 'nin daireye girişi ya da çıkışı ile ilgili olup olmadığı kaydedilir.
4. Daire büyüklüğünün yanı sıra merkezi sağa en uzak olan dairelerin içindeki olay sayıları kaydedilir. Sıralanan daire merkezleri, dairenin içine giriş ya da çıkış olması durumuna göre olayları ekleyip çıkararak aşağı doğru taranır. Her w dairesel alanı için, r yarıçap ve n olay sayısı olmak üzere gözlenen olay sayısı $\mu(w) = \int \mu dy_w = \mu\pi r^2$ olarak hesaplanır.

5. Tüm olay çiftleri için 1'den 4'e kadar bütün adımlar tekrarlanır. $(n, \mu(w))$ ikilisine göre olabilirlik hesaplanır ve en büyüğü tarama istatistiği olarak tanımlanır.
6. 1'den 5'e kadar tüm adımlar her bir Monte Carlo sayı üretimi için tekrarlanır.

Algoritma 4: Dairesel Pencere. Değişken Büyüklük. Kesikli Homojen Olmayan Poisson ya da Bernoulli Modeli

Algoritma 3'teki adımlar uygulanır. Tek farkı burada gözlenen olay sayısı belirli bir özellik gösteren olay sayısı değil gözlenen olayların tüm sayısı olarak tanımlanmaktadır. M , toplam gözlenen olay sayısı olmak üzere algoritma karmaşıklığı $O(Rn(G)^3 \log M)$ olarak bulunur.

Algoritma 5: Dairesel Pencere. Değişken Büyüklük. Kılavuz Olay Merkezli Daire. Homojen Poisson Süreci.

1. Bir kılavuz olay(grid point) belirlenir. Bu olayın diğer olaylara olan uzaklığı belirlenir ve bu uzaklıklar artan sırada sıralanır.
2. Bu kılavuz olay merkezli bir daire çizilir. Dairenin yarıçapı sürekli artırılır. n , daireye giren olay sayısı olmak üzere daireye giren her olaydan sonra daire içindeki olay sayısı kaydedilir. Daire içindeki gözlenen olay sayısı, $\mu(w) = \mu \pi r^2$ olarak hesaplanır.
3. 1. ve 2. adımlar her kılavuz nokta için tekrarlanır. $(n, \mu(w))$ ikilisine göre olabilirlik hesaplanır ve en büyüğü tarama istatistiği olarak tanımlanır.
4. İlk 3 adım her bir Monte Carlo sayı üretimi için tekrarlanır.

D , kılavuz olay sayısı olmak üzere algoritma karmaşıklığı $O(RDn(G) \log n(G))$ olarak hesaplanır.

Algoritma 6: Dairesel Pencere. Değişken Büyüklük. Kılavuz Olay Merkezli Daire. Kesikli Homojen Olmayan Poisson Süreci.

Bir önceki algoritma ile aynı şekilde uygulanır. Bu algoritma toplam gözlenen olay sayısına bağlıdır. 1. ve 2. adımların karmaşıklıkları $O(Dn(G) \log n(G))$ ve 3. adımdan 6. adıma kadar olan uygulamalardaki karmaşıklık ise $O(RDn(G))$ ile hesaplanır.

Bunu izleyen bölümde tarama istatistiklerinin geniş kullanım alanları ile ilgili bazı örneklerine yer verilmiştir.

5. BAZI KULLANIM ALANLARI

Tarama istatistikleri Naus'un 1965 yılındaki çalışmalarından sonra sürekli olarak geliştirilmiş ve birçok alanda kullanılmaya başlanmıştır. Bu alanlara örnek olarak, enfeksiyon hastalıkları, kanser, kardiyoloji, pediatri, romatoloji, nörolojik hastalıklar, karaciğer hastalıkları, şeker hastalığı, parazitoloji, kazalar, tıp, ormancılık, psikoloji, tıbbi görüntüleme, tarih, kriminoloji, astronomi ve jeoloji verilebilir. Kullanım alanları ile ilgili bazı örnekler aşağıda verilmiştir.

Medikal Görüntüleme

Medikal görüntüleme, mamografi ile tümörleri bulmaya, beyin taramaları ile kesin fiziksel ya da zihinsel aktiviteleri ölçmek için kullanılır. Bu alanda iki ve üç boyutlu uygulamalar yapılmıştır. Priebe (1998), mamografi görüntülerindeki rasgele tarama bölgelerine dayanarak yaptığı tarama istatistiklerinde göğüsteki kireçlenme kümelerini araştırmış ve tarama bölgelerini göğüs dokularını kullanarak tanımlamıştır. Worsley et al. (1992) rasgele alan taraması ile beyindeki aktivite merkezleri üzerine bir çalışma yapmışlardır.

Kanser Epidemiyolojisi

Mekansal tarama istatistikleri genellikle önemli kanser kümelerinin tespiti ve değerlendirilmesi için kullanılmaktadır. Lösemi, mekansal tarama istatistikleri ile araştırılan ilk kanser türüdür (Turnbull et al. 1990). Hjalmar et al. (1996) İsveç'teki lösemili çocuklar üzerinde bir araştırma yapmıştır.

Bulaşıcı Hastalık

Bulaşıcı hastalıkların izlenmesinde mekansal tarama istatistikleri iki farklı amaçla kullanılabilir. Bunlardan ilki geçmişe ait verilerin kullanılarak geriye doğru tarama yapılması işlemidir. Bu tür kümeler tespit edildikten sonra zaman içinde değişmeyeceği gibi farklılık da gösterebilmektedir. Bu konuyla ilgili Fevre et al. (2001), Uganda'daki uyku hastalığı üzerine mekansal tarama istatistiklerini kullanarak bir araştırma yapmışlardır. Uyku hastalığının çeçe sineği ile insanlara bulaşan bir parazitten kaynaklandığı bilinmektedir. Yapılan analizler sonucunda anlamlı bulunan kümenin hayvan pazarı yanında yaşayan insanları kapsadığı görülmüştür. Mekansal tarama istatistiklerinin bulaşıcı hastalıklar için ikinci kullanım amacı ise ileriye doğru taramalar içindir. Burada veriler gerçek zamanlı ve sürekli toplanmakta ve hızlı bir

şekilde yayılmakta olan bir hastalıkla ilgili kümelenmeler tespit edilebilmektedir. Kümeler tespit edilir edilmez hastalığa neden olan sebeplerin ortadan kaldırılması için çalışmalar yapılabilmektedir. Mostashari et al. (2003) ölü kuş sayısının günlük olarak kaydedilmesi ile Batı Nil virüsü üzerine bir çalışma yapmışlar ve önerilerde bulunmuşlardır.

Veterinerlik

Bu alanda özellikle son yıllarda İsviçre (Schwermer et al. 2002), Fransa (Abrial et al. 2003), İspanya (Allepuz et al. 2007), Hollanda (Heres et al. 2008) ve İrlanda (Sheridan et al. 2005) gibi ülkelerde mekansal tarama istatistikleri kullanılarak deli dana hastalığı üzerine yapılan çalışmalar oldukça popülerdir. Bunun yanında yabani hayat üzerine de tarama istatistikleri kullanılarak oldukça çalışma yapılmıştır. Örneğin Almanya'da tilkiler (Berke et al. 2002), İrlanda'da porsuklar (Olea et al. 2003) üzerinde etkili olan çeşitli hastalıklarla ilgili araştırmalar yapılmıştır.

Ormancılık

Tuia et al. (2008) orman yangınlarının kümelenmeleri üzerine uzay-zaman tarama istatistiklerini kullanarak bir araştırma yapmışlardır. Analizler sonucunda yangın oluşumundaki mekansal ve zamansal durumların ileriye doğru çalışmalar üzerinde etkili olacağı sonucuna varmışlardır.

Astronomi

Mekansal ve uzay-zaman tarama istatistikleri astronomi alanında kullanılabilir. Örneğin, yıldızların, galaksilerin ya da diğer gök cisimlerinin anlamlı kümeler oluşturup oluşturmadığı araştırılabilir. Yıldız oluşumu üzerine Marcos and Marcos (2008) mekansal tarama istatistiklerini kullanarak karşılıklı çekim kuvvetleri ile bir arada bulunan yıldız kümelerini araştırmışlardır ve istatistiksel olarak anlamlı bir küme tespit etmişlerdir.

Arkeoloji ve Tarih

Alt and Vach (1991), genetik olduğu bilinen diş yapısı özelliğine göre tarih öncesindeki mezarların yerlerini karşılaştırmak için mezarların konumları ile ilgili araştırma yapmışlardır. Bu çalışmalarında biyolojik olarak birbirleriyle ilişki olan yani

aynı diř yapısına sahip insanların birbirlerine yakın yerlerde gömölüp gömölmediđini amaçlamıřlardır.

Psikoloji

Margai and Henry (2003), New York'ta ilkokul çađındaki çocuklar arasında öğrenme güçlüđü ile ilgili yüksek prevalanslı kümelenmelerini arařtırmıřlar ve řehrin kuzeyinde istatistiksel olarak anlamlı bir küme tespit etmiřlerdir.

Kazalar

Nkhoma et al. (2004), kazayla zehirlenme sonucu ölümler üzerine mekansal tarama istatistiklerini kullanarak bir arařtırma yapmıřlardır. Yiannakoullias et al. (2002), yařlılarda düşme sonucu yaralanmalar üzerine bir arařtırma yapmıřlardır.

Keřif

Savařlar sona erse de mayınlar her yıl binlerce insanı yaralamaktadır. Bu nedenle mayınların yerlerinin tespit edilmesi, etkisizleştirilmeleri ve yok edilmeleri büyük önem taşımaktadır. Olası mayınların tespiti için havadan geniş bir alanda tarama yapmak mümkün olsa da, noktasal konumlar önemli olduđundan havadan yapılan taramalar her zaman dođru sonuç vermez. Tarama istatistikleri kullanılarak mayınları kapsayan en olası kümeler tespit edilebilir.

Bunun yanında maden, petrol ve uranyum yataklarının tespitinde de tarama istatistiklerinden yararlanılabilir. Conover et al. (1979) uranyum yataklarının tespiti için radyasyon ölçüm verilerini kullanarak yüksek radyasyonlu kümeleri arařtırmıřlardır.

Suç Bilimi

Beato et al. (2001), Brezilya'daki cinayetlerin cođrafi dađılımını arařtırmak için mekansal tarama istatistiklerini kullanmıřlardır. İstatistiksel olarak anlamlı bulunan küme uyuřturucu kaçakçılıđı ile bilinen bölgede ortaya çıkmıřtır (Glaz et al., 2009).

Bunu izleyen bölümde uygulama ařamasında kullanılan SaTScan programı ile ilgili bilgi verilecektir.

6. SaTScan PROGRAMI

Bu tez çalışmasındaki uygulamalar için SaTScan programdan yararlanılmıştır. Bu nedenle burada programla ilgili genel bir bilgi verilmiştir.

SaTScan programı zamansal, mekansal ve uzay-zaman analizlerini yapan ücretsiz bir programdır. Bu program www.satscan.org adresinden alınmıştır. Programın tasarlanma amaçları aşağıda verilmiştir:

1. Hastalıkların coğrafik gözetimleri için iki ve üç boyutlu hastalık kümelerinin tespit edilmesi ve bu kümelerin anlamlı olup olmadıklarını tespit etmek,
2. Bir hastalığın bir zaman, mekan veya uzay-zaman içinde rasgele dağılıp dağılmadığını tespit etmek,
3. Hastalık kümelerini istatistiksel olarak değerlendirmek,
4. Hastalık salgınlarının erken tanısı için ileriye dönük gerçek zamanlı ya da periyodik zamanlarda tarama yapmaktır.

Bu yazılım aynı zamanda, arkeoloji, astronomi, botanik, kriminoloji, ekoloji, ekonomi, mühendislik, ormancılık, genetik, coğrafya, jeoloji, tarih, nöroloji, zooloji gibi başka alanlarda da benzer problemler için kullanılabilir.

SaTScan programı Bölüm 4'te verilen Algoritma 6'yı kullanmaktadır. Bu program yardımıyla geriye doğru ve ileriye doğru tarama işlemleri Poisson, Bernoulli, normal, üstel vb. dağılımlar için yapılabilir.

Program ilk çalıştırıldığında veri girişi için açılan pencere aşağıdaki gibidir:

The screenshot shows the SaTScan software interface with the 'Input' tab selected. The window contains the following elements:

- Case File:** An empty text box with a browse button (...).
- Control File:** A text box containing '(Bernoulli Model)' with a browse button (...).
- Time Precision:** Radio buttons for 'None', 'Year' (selected), 'Month', 'Day', and 'Generic'.
- Study Period:** Two sets of date pickers. 'Start Date' is set to 2000-1-1 and 'End Date' is set to 2000-12-31.
- Population File:** A text box containing '(Poisson Model)' with a browse button (...).
- Coordinates File:** An empty text box with a browse button (...).
- Grid File:** An empty text box with a browse button (...).
- Coordinates:** Radio buttons for 'Cartesian' and 'Lat/Long' (selected).
- Advanced >>** A button at the bottom right.

Kullanılacak modele göre önceden hazırlanmış veri dosyaları programa yüklenir. Poisson modeli için populasyon dosyasının eklenmesi gerekirken, Bernoulli modelinde ise kontrol dosyası eklenmelidir. Zaman boyutunda bir analiz yapılacaksa zaman aralığı gün ay ya da yıl gibi çeşitli seçeneklerle belirlenir. Koordinatların kartezyen ya da enlem/boylam olmak üzere veri girişinde iki seçeneği bulunmaktadır.

Veri girişinden sonra istenen model ve tarama şekli için aşağıdaki pencerede gerekli işaretlemeler yapılabilir.

The screenshot shows the 'Analysis' window of a software application. The window has three tabs: 'Input', 'Analysis', and 'Output'. The 'Analysis' tab is active. It contains several sections: 'Type of Analysis' with 'Retrospective Analyses' (Purely Spatial, Purely Temporal, Space-Time, Spatial Variation in Temporal Trends) and 'Prospective Analyses' (Purely Temporal, Space-Time). 'Probability Model' includes 'Discrete Scan Statistics' (Poisson, Bernoulli, Space-Time Permutation, Multinomial, Ordinal, Exponential, Normal) and 'Continuous Scan Statistics' (Poisson, ...). 'Scan For Areas With:' has options for High Rates, Low Rates, and High or Low Rates. 'Time Aggregation' has 'Units' (Year, Month, Day) and 'Length' (1 Years). An 'Advanced >>' button is at the bottom right.

Bu pencerede istenen model için gerekli seçenekler işaretlenebilir. Yüksek, düşük ya da her iki oranlı kümenin de aranması yine bu pencerede işaretlenecek seçeneğe göre yapılır. Eğer zaman boyutunda bir analiz yapılıyorsa zamanın değişim bilgisi de burada verilmektedir.

The screenshot shows the 'Output' window of a software application. The window has three tabs: 'Input', 'Analysis', and 'Output'. The 'Output' tab is active. It contains a 'Results File:' field with a text box and a button. Below it is a section for 'Additional Output Files' with a table of checkboxes for 'ASCII' and 'dBase' formats. The table has five rows: Cluster Information, Stratified Cluster Information, Location Information, Risk Estimates for Each Location, and Simulated Log Likelihood Ratios/Test Statistics. An 'Advanced >>' button is at the bottom right.

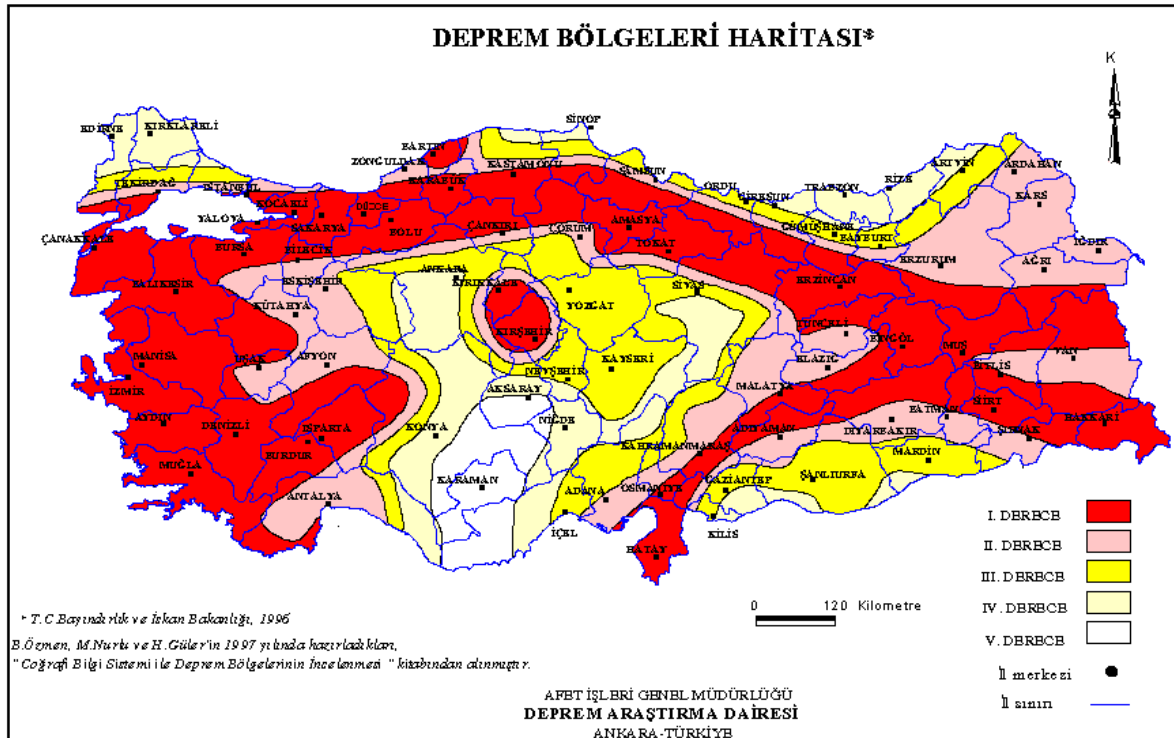
Açılan son pencerede ise, sonuçların kaydedileceği yer ve istenen sonuçlar için birden fazla işaretlenebilecek seçenekler yer almaktadır.

Bunu izleyen bölümde uygulamada kullanılacak olan deprem verileri ile ilgili genel bilgi verilmiştir.

7. DEPREM OLGUSU ve BAZI TANIMLAR

Yerkabuğu içindeki kırılmalar nedeniyle ani olarak ortaya çıkan titreşimlerin dalgalar halinde yayılarak geçtikleri ortamları ve yer yüzeyini sarsma olayına "DEPREM" denir.

Bilindiği gibi yurdumuz dünyanın en etkin deprem kuşaklarından birinin üzerinde bulunmaktadır. Deprem Bölgeleri Haritası'na göre, yurdumuzun %92'sinin deprem bölgeleri içerisinde olduğu, nüfusumuzun %95'inin deprem tehlikesi altında yaşadığı ve ayrıca büyük sanayi merkezlerinin %98'i ve barajlarımızın %93'ünün deprem bölgesinde bulunduğu bilinmektedir. Son 58 yıl içerisinde depremlerden, 58.202 vatandaşımız hayatını kaybetmiş, 122.096 kişi yaralanmış ve yaklaşık olarak 411.465 bina yıkılmış veya ağır hasar görmüştür. Sonuç olarak denilebilir ki, depremlerden bir yılda ortalama 1.003 vatandaşımızın öldüğünü ve 7.094 bina yıkıldığını söyleyebiliriz. (www.deprem.gov.tr)



Resim 7.1. Türkiye Deprem Bölgeleri Haritası

DEPREM PARAMETRELERİ

Odak Noktası (Hiposantr)

Yerin içinde depremin enerjisinin ortaya çıktığı noktadır. Bu noktaya odak noktası veya iç merkez de denir. Gerçekte, enerjinin ortaya çıktığı bir nokta olmayıp bir alandır, fakat pratik uygulamalarda nokta olarak kabul edilmektedir.

Dış Merkez (Episantr)

Odak noktasına en yakın olan yer üzerindeki noktadır. Burası aynı zamanda depremin en çok hasar yaptığı veya en kuvvetli olarak hissedildiği noktadır. Depremin dış merkez alanı depremin şiddetine bağlı olarak çeşitli büyüklüklerde olabilir. Bazen büyük bir depremin odak noktasının boyutları yüzlerce kilometreyle de belirlenebilir.

Odak Derinliği

Depremde enerjinin açığa çıktığı noktanın yeryüzünden en kısa uzaklığı, depremin odak derinliği olarak adlandırılır. Depremler odak derinliklerine göre sınıflandırılabilir. Yerin 0-60 km. derinliğinde olan depremler sığ deprem olarak nitelenir. Yerin 70-300 km. derinliklerinde olan depremler orta derinlikte olan depremlerdir. Derin depremler ise yerin 300 km.den fazla derinliğinde olan depremlerdir. Türkiye'de olan depremler genellikle sığ depremlerdir. Derin depremler çok geniş alanlarda hissedilir, buna karşılık yaptıkları hasar azdır. Sığ depremler ise dar bir alanda hissedilirken bu alan içinde çok büyük hasar yapabilirler.

Eş Şiddet (İzoseit) Eğrileri

Aynı şiddetle sarsılan noktaları birbirine bağlayan noktalara denir. Bunun tamamlanmasıyla eş şiddet haritası ortaya çıkar. Genelde kabul edilmiş duruma göre, eğrilerin oluşturduğu yani iki eğri arasında kalan alan, depremlerden etkilenme yönüyle, şiddet bakımından sınırlandırılmış olur.

Şiddet

Herhangi bir derinlikte olan depremin, yeryüzünde hissedildiği bir noktadaki etkisinin ölçüsü olarak tanımlanmaktadır. Diğer bir deyişle depremin şiddeti, onun yapılar,

doğa ve insanlar üzerindeki etkilerinin bir ölçüsüdür. Bu etki, depremin büyüklüğü, odak derinliği, uzaklığı, yapıların depreme karşı gösterdiği dayanıklılığa göre farklılık göstermektedir. Şiddet depremin kaynağındaki büyüklüğü hakkında doğru bilgi vermemekle beraber, deprem dolayısıyla oluşan hasarı yukarıda belirtilen etkenlere bağlı olarak yansıtır.

Magnitüd

Deprem sırasında açığa çıkan enerjinin bir ölçüsü olarak tanımlanmaktadır. Enerjinin doğrudan doğruya ölçülmesi olanağı olmadığından, Amerika Birleşik Devletleri'nden Prof.C.Richter tarafından 1930 yıllarında bulunan bir yöntemle depremlerin aletsel bir ölçüsü olan "Magnitüd" tanımlanmıştır. Prof.Richter, episantrdan 100 km. uzaklıkta ve sert zemine yerleştirilmiş özel bir sismografla (2800 büyütme, özel periyodu 0,8 saniye ve %80 sönümü olan bir Wood-Anderson torsiyon Sismografı ile) kaydedilmiş zemin hareketinin mikron cinsinden (1 mikron 1/1000 mm) ölçülen maksimum genliğinin 10 tabanına göre logaritmasını bir depremin "magnitüdü" olarak tanımlamıştır. Bugüne dek tüm dünyada meydana gelen depremler istatistik olarak incelendiğinde kaydedilen en büyük magnitüd değerinin 8,9 olduğu görülmektedir(31 Ocak 1906 Colombiya-Ekvator ve 2Mart 1933 Sanriku-Japonya depremleri). Ülkemizde ise kaydedilen en büyük magnitüd değer, 26-27 Aralık 1939'da Erzincan'da meydana gelen depremde kaydedilmiştir. Şiddeti Richter ölçeğine göre 8 olan deprem sonucunda toplam 32.962 kişi ölmüş, yaklaşık 100.000 kişi yaralanmış ve 116.720 bina yıkılmıştır. Dünyanın büyük depremleri arasında sayılan bu deprem Türkiye'nin en ciddi deprem felaketlerinden biridir.

Tanım	Büüklük	Yıllık Ortalama
Çok Çok Şiddetli (Great)	8 >=	1
Çok Şiddetli (Major)	7 – 7,9	18
Şiddetli (Strong)	6 – 6,9	120
Orta Şiddette (Moderate)	5 – 5,9	800
Hafif (Light)	4 – 4,9	6,200 (tahmini)
Çok Hafif (Minor)	3 – 3,9	49,000 (tahmini)
Çok Çok Hafif (Very Minor)	< 3,0	Magnitüd 2 - 3: günde yaklaşık 1,000 Magnitüd 1 - 2: günde yaklaşık 8,000

Tablo 7.1. Deprem Büyüklükleri ve Dünyada 1 Yılda Ortalama Meydana Geliş Sayıları

Bunu izleyen bölümde, ülkemizdeki coğrafik ve jeolojik koşullar nedeniyle yaşamımızın bir parçası olan deprem verileri tez çalışmasında kullanılmış, tek, iki ve üç boyutlu tarama istatistikleri kullanılarak çeşitli dağılımlar altında kümelenme olup olmadığı incelenmiş ve elde edilen sonuçlar değerlendirilmiştir.

8. UYGULAMALAR

Türkiye'deki deprem verileri üzerine tek, iki ve üç boyutlu tarama istatistikleri uygulanmış ve çeşitli sonuçlar elde edilmiştir. Tek boyutlu tarama istatistikleri uygulamasında aylık ve günlük verilerin zamansal değişimi incelenmiştir. Aylık deprem verileri 2000-2011 yılları arasında meydana gelen depremlerin Kandilli Rasathanesi'nin internet sitesindeki düzenlenmiş aylık bültenlerden alınmıştır. Günlük deprem verileri, Deprem Dairesi Başkanlığı'nın internet sitesinden son 3 yılın verileri alınarak düzenlenmiştir. Aynı veriler iki ve üç boyutlu analizlerde düzenlenerek kullanılmıştır.

Analizler için SaTScan programı kullanılmış; bu programda maksimum küme genişliği tüm verilerin en fazla %50'sini kapsayacak şekilde belirlenmekte, p değerinin hesaplanabilmesi için Monte Carlo Simülasyon Yöntemi kullanılmaktadır.

8.1 TEK BOYUTLU TARAMA İSTATİSTİKLERİ

8.1.1 Aylık Deprem Verileri

Geriye Doğru Tarama

1. Türkiye'de meydana gelen depremlerin zamansal değişimlerini incelenmek üzere 2000–2011 yılları arasında her ay meydana gelen deprem sayısı dikkate alınarak tek boyutlu tarama istatistiği ile analiz yapılmıştır. Analizde aylık deprem sayılarının Poisson dağıldığı varsayımı altında yüksek oranlı kümeler araştırılmıştır. 2000 yılının 1. ayı ile 2011 yılının 12. ayı arasında her ay meydana gelen depremlerin toplam sayısı 82498'dir. Bu analizde ilgilenilen hipotezler aşağıdaki gibidir:

H_0 : 2000-2011 yılları arasında meydana gelen depremlerde zaman içinde bir kümelenme yoktur.

H_1 : 2000-2011 yılları arasında meydana gelen depremlerde zaman içinde bir kümelenme vardır.

Analiz sonucunda 2007 yılının 12. ayı ile 2011 yılının 12. ayı arasındaki zaman aralığı anlamlı bir küme oluşturmuştur ($p = 0,001 < \alpha = 0,05$). Bu durumda bu zaman aralığında Türkiye'deki deprem sayılarında artma olduğunu söyleyebiliriz. 2007 yılının 12. ayına kadar meydana gelen depremlerin aylık ortalaması 372,5789 iken, kümelenmenin olduğu zaman aralığındaki depremlerin aylık ortalaması ise 961,2857 olarak bulunmuştur. Analizle ilgili program çıktısı Ek 1'de verilmiştir.

2. 2000–2011 yılları arasında her ay meydana gelen 4'ten büyük deprem sayılarının Poisson dağıldığı varsayımı altında tek boyutlu tarama istatistikleri uygulanmıştır. Bu zaman aralığında meydana gelen 4'ten büyük depremlerin toplam sayısı 1877'dir.

H_0 : 2000-2011 yılları arasında meydana gelen 4'ten büyük depremlerde zaman içinde bir kümelenme yoktur.

H_1 : 2000-2011 yılları arasında meydana gelen 4'ten büyük depremlerde zaman içinde bir kümelenme vardır.

Analiz sonucunda 2011 yılının 10. ve 11. ayları arasında meydana gelen 4'ten büyük depremlerin yoğunlaştığı görülmektedir. Bu zaman aralığında oluşan kümelenme anlamlı bulunmuştur ($p = 0,001 < \alpha = 0,05$). Kümelenme dışındaki aylık ortalama deprem 11,9647 olurken, kümelenme içindeki ortalama ise 89 olarak bulunmuştur. Analizle ilgili program çıktısı Ek 2'de verilmiştir.

İleriye Doğru Tarama

1. 2000-2011 yılları arasındaki aylık deprem verileri kullanılarak zaman boyutunda ileriye dönük tahminler yapılmıştır. Deprem sayılarının Poisson dağıldığı varsayımı altında kurulan hipotezler aşağıda verilmiştir:

H_0 : 2000-2011 yılları arasında meydana gelen depremlerde zaman içinde bir kümelenme yoktur.

H_1 : 2000-2011 yılları arasında meydana gelen depremlerde zaman boyutunda bir kümelenme vardır.

Analiz sonucuna göre 10 yıllık bir zamanda meydana gelen 82498 depremden 47103'ü 2007 yılının son ayı ile 2011 yılının son ayı arasında oluşmuştur. Oluşan bu kümenin tekrarlanma periyodu 83 yıldır. Analizle ilgili program çıktısı Ek 3'te verilmiştir.

2. 2000-2011 yılları arasındaki 4'ten büyük aylık deprem verileri kullanılarak zaman boyutunda ileriye dönük tahminler yapılmıştır. Deprem sayılarının Poisson dağıldığı varsayımı altında yapılan analiz sonucu aşağıda verilmiştir:

H_0 : 2000-2011 yılları arasında meydana gelen 4'ten büyük depremlerde zaman boyutunda bir kümelenme yoktur.

H_1 : 2000-2011 yılları arasında meydana gelen 4'ten büyük depremlerde zaman boyutunda bir kümelenme vardır.

Analiz sonucuna göre 10 yıllık bir zamanda meydana gelen 1877 depremden 202'si 2011 yılının 10. ayı ile 2011 yılının 12. ayında bir kümelenme oluşturmuştur. Oluşan bu kümenin tekrarlanma periyodu 83 yıldır. Analizle ilgili program çıktısı Ek 4'te verilmiştir.

8.1.2 Günlük Deprem Verileri

Geriye Doğru Tarama

2009, 2010, 2011 yıllarındaki günlük deprem verileri kullanılarak zamansal değişim incelenmiş ve aşağıdaki sonuçlar elde edilmiştir:

Yıl	Toplam Deprem Sayısı	En Olası Küme					
		Zaman Aralığı	Meydana Gelen Deprem Sayısı	Beklenen Deprem Sayısı	Görel Risk	Olabilirlik Oranı	p-değeri
2009	15177	17.02.2009 04.03.2009	1276	665,29	2,00	233,3325	0,001
2010	18697	01.11.2010 06.11.2010	573	307,35	1,89	93,2007	0,001
2011	29577	23.10.2011 23.11.2011	6540	2593,05	2,95	2407,0805	0,001

Çizelge 8.1. 2009-2010-2011 Deprem Verileri (Günlük) ve Geriye Doğru Kümelenmeler

H_0 : 2009 yılında meydana gelen depremlerde (günlük) zaman boyutunda bir kümelenme yoktur.

H_1 : 2009 yılında meydana gelen depremlerde (günlük) zaman boyutunda bir kümelenme vardır.

2009 yılına ait günlük deprem verilerine göre, en olası küme 17.02.2009 ile 04.03.2009 zaman aralığını kapsamaktadır. Bu kümedeki deprem sayısı 1276 olup beklenen deprem sayısı 665,29 olarak hesaplanmıştır. Olabilirlik oranı 233,3325 ile oluşan bu küme anlamlı bulunmuştur ($p = 0,001 < \alpha = 0,05$) (Program çıktısı Ek 5'te verilmiştir).

H_0 : 2010 yılında meydana gelen depremlerde zaman boyutunda bir kümelenme yoktur.

H_1 : 2010 yılında meydana gelen depremlerde zaman boyutunda bir kümelenme vardır.

2010 yılındaki deprem verilerine göre, en olası küme 01.11.2010 ile 06.11.2010 tarihleri arasında oluşmuştur. Bu kümedeki deprem sayısı 573 olup beklenen deprem sayısı ise 307,35'dir. En olası küme 3,2007 olabilirlik oranı ile anlamlı bulunmuştur ($p = 0,001 < \alpha = 0,05$). Analizle ilgili program çıktısı Ek 6'da verilmiştir.

H_0 : 2011 yılında meydana gelen depremlerde zaman boyutunda bir kümelenme yoktur.

H_1 : 2011 yılında meydana gelen depremlerde zaman boyutunda bir kümelenme vardır.

2011 yılındaki verilere göre, en olası küme 23.10.2011 ile 23.11.2011 tarihleri arasında oluşmuştur. Bu kümedeki deprem sayısı 6540 olup beklenen deprem sayısı ise 2593,05 olarak bulunmuştur. Bu küme 2407,0805 olabilirlik oranı ile anlamlı bulunmuştur ($p = 0,001 < \alpha = 0,05$). Analizle ilgili program çıktısı Ek 7'de verilmiştir.

İleriye Doğru Tarama

2009, 2010, 2011 yıllarındaki günlük deprem verilerine ileriye dönük tarama istatistikleri uygulanarak tahminler yapılmış ve karşılaştırmalar aşağıda gösterilmiştir:

Yıl	Toplam Deprem Sayısı	En Olası Küme						
		Zaman Aralığı	Meydana Gelen Deprem Sayısı	Beklenen Deprem Sayısı	Görelî Risk	Olabilirlik Oranı	p-değeri	Tekrar Periyodu
2009	15177	22.12.2009 31.12.2009	494	415,81	1,19	7,1375	0,004	250 gün
2010	18697	12.07.2010 31.12.2010	9103	8861,87	1,05	6,2317	0,007	143 gün
2011	29577	23.10.2011 31.12.2011	10436	5672,30	2,30	2108,5371	0,001	2,7 yıl

Çizelge 8.2. 2009-2010-2011 Deprem Verileri (Günlük) ve İleriye Doğru Kümelenmeler

H_0 : 2009 yılında meydana gelen depremlerde zaman boyutunda bir kümelenme yoktur.

H_1 : 2009 yılında meydana gelen depremlerde zaman boyutunda bir kümelenme vardır.

2009 yılına ait günlük deprem verilerine göre, en olası küme 22.12.2009 ile 31.12.2009 zaman aralığını kapsamaktadır. Bu kümedeki deprem sayısı 494 olup beklenen deprem sayısı 415,81 olarak hesaplanmıştır. Görelî risk 1,19 olarak bulunmuştur. Olabilirlik oranı 7,1375 ile oluşan bu küme anlamlıdır ($p = 0,004 < \alpha = 0,05$). Bu kümelenmenin tekrarlanma periyodu 250 gündür. Analizle ilgili program çıktısı Ek 8'de verilmiştir.

H_0 : 2010 yılında meydana gelen depremlerde zaman boyutunda bir kümelenme yoktur.

H_1 : 2010 yılında meydana gelen depremlerde zaman boyutunda bir kümelenme vardır.

2010 yılındaki deprem verilerine göre, en olası küme 12.07.2010 ile 31.12.2010 tarihleri arasında oluşmuştur. Bu kümedeki deprem sayısı 9103 olup beklenen deprem sayısı ise 8861,87'dir. En olası küme 3,2007 olabilirlik oranı ile anlamlıdır ($p = 0,007 < \alpha = 0,05$). Bu kümelenmenin tekrarlanma periyodu 143 gündür. Analizle ilgili program çıktısı Ek 9'da verilmiştir.

H_0 : 2011 yılında meydana gelen depremlerde zaman boyutunda bir kümelenme yoktur.

H_1 : 2011 yılında meydana gelen depremlerde zaman boyutunda bir kümelenme vardır.

2011 yılındaki verilere göre, en olası küme 23.10.2011 ile 31.12.2011 tarihleri arasında oluşmuştur. Bu kümedeki deprem sayısı 10436 olup beklenen deprem sayısı ise 5672,30 olarak bulunmuştur. Göreli risk 2,30 olarak hesaplanmış olup bu küme 2407,0805 olabilirlik oranı ile anlamlıdır ($p = 0,001 < \alpha = 0,05$). Bu kümelenmenin tekrarlanma periyodu 2,7 yıldır. Analizle ilgili program çıktısı Ek 10'da verilmiştir.

8.2 İKİ BOYUTLU TARAMA İSTATİSTİKLERİ

$36^\circ - 42^\circ$ enlemleri ile $26^\circ - 45^\circ$ boylamları arasında yer alan ülkemizde meydana gelen depremler iki boyutlu olarak incelenirken bu enlem-boylam sınırları arasındaki depremler analizde yer almıştır.

1. 2009 yılında meydana gelen depremler üzerinde Bernoulli modeli kullanılarak mekansal analiz yapılmıştır. Burada 4'ten büyük depremler 1 ile gösterilirken, 4'ten küçük depremler 0 olarak tanımlanmıştır. 2009 yılında toplam 13865 deprem meydana gelirken bu depremlerin 47'si 4'ten büyüktür.

H_0 : 2009 yılında meydana gelen depremlerde mekansal boyutta bir kümelenme yoktur.

H_1 : 2009 yılında meydana gelen depremlerde mekansal boyutta bir kümelenme vardır.

Analiz sonucuna göre, olabilirlik oranı en yüksek olan küme en olası küme olarak belirlenmiş ve bu küme istatistiksel olarak anlamlı bulunmuştur ($p = 0,0044 < \alpha = 0,05$). Analizle ilgili program çıktısı Ek 11'de verilmiştir.

2010 yılında meydana gelen depremler üzerinde Bernoulli modeli kullanılarak mekansal analiz yapılmıştır. Burada 4'ten büyük depremler 1 ile gösterilirken, 4'ten küçük depremler 0 olarak tanımlanmıştır. 2010 yılında toplam 17164 deprem meydana gelirken bu depremlerin 74'ü 4'ten büyüktür.

H_0 : 2010 yılında meydana gelen günlük depremlerde mekansal boyutta bir kümelenme yoktur.

H_1 : 2010 yılında meydana gelen günlük depremlerde mekansal boyutta bir kümelenme vardır.

Analiz sonucuna göre, olabilirlik oranı en yüksek olan küme en olası küme olarak belirlenmiş ve tüm kümeler istatistiksel olarak anlamlı bulunmuştur ($p = 41 \times 10^{-10} < \alpha = 0,05$). Analizle ilgili program çıktısı Ek 12’de verilmiştir.

2011 yılında meydana gelen depremler üzerinde Bernoulli modeli kullanılarak mekansal analiz yapılmıştır. Burada 4’ten büyük depremler 1 ile gösterilirken, 4’ten küçük depremler 0 olarak tanımlanmıştır. 2011 yılında toplam 28213 deprem meydana gelirken bu depremlerin 272’si 4’ten büyüktür.

H_0 : 2011 yılında meydana gelen günlük depremlerde mekansal boyutta bir kümelenme yoktur.

H_1 : 2011 yılında meydana gelen günlük depremlerde mekansal boyutta bir kümelenme vardır.

Analiz sonucuna göre, olabilirlik oranı en yüksek olan küme en olası küme olarak belirlenmiş ve tüm kümeler istatistiksel olarak anlamlı bulunmuştur ($p = 0,000 < \alpha = 0,05$). Analizle ilgili program çıktısı Ek 13’te verilmiştir.

En Olası Küme							
Yıl	Toplam Deprem Sayısı	4’ten Büyük Deprem Sayısı	4’ten Büyük Beklenen Deprem Sayısı	Görel Risk	Olabilirlik Oranı	p-değeri	Koordinatlar-Yarıçap
2009	58	6	0,20	34,84	15,3848	0,0044	(36,4603 E; 27,0420 B) 47,67 km
2010	215	16	0,93	21,75	32,7266	41×10^{-10}	(38,7690 E; 40,0363 B) 6,23 km
2011	7743	187	74,65	5,82	101,2620	$< 10^{-18}$	(37,3710 E; 42,7737 B) 189,78 km

Çizelge 8.3. Mekansal Tarama ile Bulunan En Olası Kümeler

2.Küme							
Yıl	Toplam Deprem Sayısı	4'ten Büyük Deprem Sayısı	4'ten Büyük Beklenen Deprem Sayısı	Görel Risk	Olabilirlik Oranı	p-değeri	Koordinatlar-Yarıçap
2010	306	11	1,32	9,62	14,4818	0,0072	(36,0008 E; 27,2427 B) 112,86 km

Çizelge 8.4. Mekansal Tarama ile Bulunan İkincil Olası Kümeler

Burada bulunan anlamlı kümeler Resim 8.1'de verilen haritada gösterilmiştir.



Resim 8.1. Mekansal Taramada Kümelenmeler

Mavi ile gösterilen daire 2009 yılındaki anlamlı en olası kümelenmeyi göstermektedir. Kümelenme Ege Denizi'nde meydana gelmiş olup kıyıya çok yakın olduğundan Datça ile Bodrum kıyılarını da etkilemektedir.

Kırmızı ile gösterilen daire 2010 yılındaki anlamlı en olası kümelemeyi göstermektedir. Bu kümenin merkezi Elazığ olup, ilçeleri ile Tunceli ve Bingöl illerinin de Elazığ'a sınır bölgelerini kapsamaktadır.

Kesikli kırmızı ile gösterilen daire 2010 yılındaki 2. anlamlı kümelenmeyi göstermektedir. Bu kümelenme Ege Denizi'nde meydana gelmiş olup Muğla kıyılarını da etkilemektedir.

Sarı ile gösterilen daire 2011 yılındaki anlamlı en olası kümelenmeyi göstermektedir. Bu kümenin merkezi Şırnak şehri üzerinde bulunmaktadır. Bu küme Şırnak (457.997 kişi), Batman (524.499), Mardin (764.033), Hakkari (272.165), Van (1.022.532), Siirt (310.468), Bitlis (336.624), Diyarbakır (1.570.943) şehirlerini kapsamakta ve bu küme içerisinde meydana gelen depremlerin etkili olduğu toplam nüfus 5.259.261'dir.

8.3. ÜÇ BOYUTLU TARAMA İSTATİSTİKLERİ

Veri sayısının fazla olması nedeniyle SaTScan programı analiz için yetersiz kalmıştır. Bu nedenle 2011 yılındaki deprem verileri mevsimlere göre düzenlenerek Bernoulli modeline göre üç boyutlu analiz yapılmıştır. Magnitüdü 3 ile 4 arasında olan depremler 0 ile 4'ten büyük olan depremler ise 1 olarak tanımlanmıştır. Boyutlar enlem, boylam ve zamandan oluşmaktadır.

Kış Mevsimi Geriye Doğru Tarama

H_0 : 2011 yılının kış mevsiminde meydana gelen günlük depremlerde uzay- zaman boyutunda bir kümelenme yoktur.

H_1 : 2011 yılının kış mevsiminde meydana gelen günlük depremlerde uzay- zaman boyutunda bir kümelenme vardır.

Analiz sonucunda anlamlı bir küme bulunamamıştır ($p = 0,162 > \alpha = 0,05$). Bu durumda 2011 yılında kış mevsiminde 3'ten büyük depremler için uzay-zaman

boyutunda bir kümelenme olmadığını söyleyebiliriz. Analizle ilgili program çıktısı Ek 14'te verilmiştir.

İlkbahar Mevsimi Geriye Doğru Tarama

H_0 : 2011 yılının ilkbahar mevsiminde meydana gelen günlük depremlerde uzay-zaman boyutunda bir kümelenme yoktur.

H_1 : 2011 yılının ilkbahar mevsiminde meydana gelen günlük depremlerde uzay-zaman boyutunda bir kümelenme vardır.

Analiz sonucunda anlamlı bir küme bulunamamıştır ($p = 0,084 > \alpha = 0,05$). Bu durumda 2011 yılında ilkbahar mevsiminde 3'ten büyük depremler için uzay-zaman boyutunda bir kümelenme olmadığını söyleyebiliriz. Analizle ilgili program çıktısı Ek 15'te verilmiştir.

Yaz Mevsimi Geriye Doğru Tarama

H_0 : 2011 yılının yaz mevsiminde meydana gelen günlük depremlerde uzay-zaman boyutunda bir kümelenme yoktur.

H_1 : 2011 yılının yaz mevsiminde meydana gelen günlük depremlerde uzay-zaman boyutunda bir kümelenme vardır.

Analiz sonucunda anlamlı bir küme bulunamamıştır ($p = 0,91 > \alpha = 0,05$). Bu durumda 2011 yılında yaz mevsiminde 3'ten büyük depremler için uzay-zaman boyutunda bir kümelenme olmadığını söyleyebiliriz. Analizle ilgili program çıktısı Ek 16'da verilmiştir.

Sonbahar Mevsimi Geriye Doğru Tarama

H_0 : 2011 yılının sonbahar mevsiminde meydana gelen günlük depremlerde uzay-zaman boyutunda bir kümelenme yoktur.

H_1 : 2011 yılının sonbahar mevsiminde meydana gelen günlük depremlerde uzay-zaman boyutunda bir kümelenme vardır.

Analiz sonucunda en olası küme istatistiksel olarak anlamlı bulunmuştur ($p = 0,000 < \alpha = 0,05$). Bu durumda 2011 yılında sonbahar mevsiminde 3'ten büyük

depremler için uzay-zaman boyutunda bir kümelenme olduğunu söyleyebiliriz. Analizle ilgili program çıktısı Ek 17’de verilmiştir.

2011 yılı								
Mevsim	Toplam Deprem Sayısı	Zaman Aralığı	4'ten Büyük Deprem Sayısı	4'ten Büyük Beklenen Deprem Sayısı	Görel Risk	Olabilirlik Oranı	p-değeri	Koordinatlar Yarıçap
Kış	8	22.02.2011 22.02.2011	5	0,38	15,58	10,5135	0,162	(38,8815 E; 42,5665 B) 22,84 km
İlkbahar	14	19.05.2011 19.05.2011	6	0,52	13,78	10,9837	0,084	(40,7993 E; 28,0292 B) 206,30 km
Yaz	2	27.06.2011 13.07.2011	2	0,070	31,55	6,8094	0,91	(39,1205 E; 29,0433 B) 0,95 km
Sonbahar	427	23.10.2011 26.10.2011	72	29,62	3,53	31,8002	46×10^{-9}	(38,6200 E; 43,3418 B) 22,11 km

Çizelge 8.5. Uzay-Zaman Taraması ile Bulunan En Olası Kümeler

Burada bulunan anlamlı kümeler Resim 8.2’de verilen haritada gösterilmiştir.



Resim 8.2. 2011 yılı Sonbahar Mevsiminde Bulunan Anlamalı Küme

2011 yılının sonbaharında meydana gelen k melenmenin merkez noktası Van Őehri  zerindedir. Bu k melenme Őehrin il eleri ile Van G l 'n  de kapsamaktadır. İl eleriyle birlikte Van'da toplam 1.022.532 kiŐi bu b lgede meydana gelen depremlerden etkilenmektedir.

9. SONUÇLAR

Bu çalışmada, özellikle bilgisayar teknolojilerindeki gelişmelere dayalı olarak, mekansal verilerin toplandığı ve bu verilerin analizinde sıklıkla kullanılan tarama istatistiklerinin ortaya çıktığı tarihten günümüze kadar olan gelişim süreci incelenmiş, tek, iki ve üç boyutlu tarama istatistiklerinin temel tanım ve kavramları verilmiştir. İncelenen dağılım modelleri Türkiye'deki deprem verilerine uygulanarak çeşitli sonuçlar elde edilmiştir.

Zamansal tarama istatistikleri 2000-2011 yılları arasında meydana gelen deprem verileri üzerinde uygulanmıştır. İleriye ve geriye doğru yapılan taramalarda depremlerin son yıllarda yığıldığı ortaya çıkmıştır.

Mekansal tarama istatistikleri ile yapılan çalışmada ise 2009, 2010 ve 2011 yıllarında meydana gelen deprem verileri kullanılmış ve 4'ten büyük depremlerin yoğunlaştığı bölgeler araştırılmıştır. Analiz sonucunda meydana gelen kümelenmeler karşılaştırıldığında 2009 yılında meydana gelen kümelenmenin Ege Denizi'nde meydana geldiği ve kıyıları da etkilediği görülmüştür. 2010 yılında 2 anlamlı kümelenme bulunmuştur. En olası küme Van ve çevresinde, ikinci olası küme ise 2009 yılında meydana gelen küme ile çakışacak şekilde Ege Denizi'nde tespit edilmiştir. 2011 yılında meydana gelen kümelenme ise Güneydoğu Anadolu bölgesindeki Şırnak, Hakkari, Batman, Mardin, Van, Bitlis, Diyarbakır şehirlerini kapsayacak şekilde 189,78 kilometrelik büyük bir alanda meydana gelmiştir.

Uzay-zaman tarama istatistiği kullanılarak yapılan analizde ise veri sayısının fazlalığından dolayı 2011 yılı mevsimlere bölünmüştür. Bunun sonucunda ise yalnız sonbahar mevsiminde merkezi Van şehri olmak üzere 22,11 kilometrelik bir alanda kümelenme meydana gelmiştir.

Yapılan mekansal ve uzay-zaman boyutundaki analizlere göre oluşan kümeler göz önüne alındığında, 4'ten büyük depremlerin yoğunlaştığı coğrafi konumların, şehir planlamasına, mevcut olan binaların depreme karşı dayanıklı hale getirilmesine ve acil yardım merkezlerinin bu bölgelere yakın olacak şekilde yapılandırılmasına kadar pek çok konuda stratejik bir önemi vardır.

Çalıřmada da görüldüğü üzere tarama istatistiklerinin geniş bir uygulama alanı olmasına karşın, bu konu ülkemizde yeterince bilinmemektedir. Tarama istatistiklerinin hem teorik hem de uygulama açısından geliştirilmeye oldukça açık olması nedeniyle ileriki çalışmalarda Bayesci tarama istatistikleri, çok değişkenli tarama istatistikleri vb. yönleri araştırılabilir.

KAYNAKLAR

- Alm, S. E., 1998, On the distributions of scan statistics for poisson process in two and three dimensions, *Extreme*, Vol.1, p.111-126.
- Alm, S. E., 1997, On the distributions of scan statistics of a two dimensional poisson process. *Applied Probability Trust*, Vol.29, No.1, p.1-18.
- Burnside, W.,1936, *Theory of probability*. 148 p., Cambridge.
- Cressie, N. and Chan, N. H., 1989, Spatial modeling of regional variables. *Journal of the American Statistical Association* Vol.84, p.393–401.
- Glaz, J. and Naus J., 1983, Multiple clusters on the line, *Communications in Statistics:Theory and Methods* Vol.12, p.1961-1986.
- Glaz, J., Naus, J. and Wallentein, S., 2001. *Scan Statistics*. Springer-Verlag Inc., 367 p., New York.
- Glaz J., Pozdnyakov V., Wallenstein S., 2009, *Scan Statistics Methods and Applications*, 410 p., Boston.
- Han, S., Ishioka F. and Kurihara K., 2008, Detection of hotspot for Korea earthquake data using echelon analysis and seismic wave energy, *Journal of the Faculty of Environmental Science and Technology,Okayama*, Vol.13 No.1, p.51-56.
- Hjalmar, U., Kulldorff, M., Gustafsson G., Nagarwalla N.,1996, Childhood leukemia in Sweden: Using GIS and a spatial scan statistic for cluster detection. *Statistics in Medicine*, Vol.15, p.707–715.
- Huang L., Kulldorff M., Gregorio D., 2007, A spatial scan statistic for survival data, *Biometrics*, Vol.63, p.109-118.
- Jack, R. H., Gulliford, M. C., Ferguson, J., and Moller, H., 2003, Geographical inequalities in lung cancer management and survival in South East England: evidence of variation in access to oncology services?, *British Journal of Cancer*, Vol.88, No.7, p.1025-1031.
- Jung I., Kulldorff M., Richard OJ., 2010, A spatial scan statistic for multinomial data, *Statistics in Medicine*.
- Kahveci M., 2002, Simülasyon Modelinin Monte-Carlo Simülasyon Tekniğiyle Stokastik Süreçlerde Uygulanması, *Mali Çözüm Dergisi*,Yıl.12,Sayı.59

- Karjalainen, S., 1990, Geographical variation in cancer patient survival in Finland: chance, confounding, or effect of treatment?, *Journal of Epidemiology and Community Health*, Vol.11, No.10, p.960-963.
- Kulldorff, M., Huang, L., and Konty, K., 2009, A scan statistic for continuous data based on the normal probability model. *International Journal of Health Geographics*, Vol.8, p.58.
- Kulldorff, M., Mostashari, F., Duczmal, L., Yih, K., Kleinman, K. and Platt, R., 2007, Multivariate spatial scan statistics for disease surveillance. *Statistics in Medicine*, Vol.26, p.1824–1833.
- Kulldorff, M., Heffernan, R., Hartman, J., Assunção, RM. and Mostashari, F., 2005, A space-time permutation scan statistic for the early detection of disease outbreaks. *PLoS Medicine*, Vol.2, p.216–224.
- Kulldorff, M., 1999, Spatial scan statistics: Models, calculations and applications. In Balakrishnan and Glaz (eds), *Recent Advances on Scan Statistics and Applications*. Boston, USA: Birkhäuser.
- Kulldorff, M., 1997, A spatial scan statistic. *Communications in Statistics: Theory and Methods*, Vol.26, p.1481–1496.
- Kulldorff, M. and Nagarwalla, N., 1995, Spatial disease clusters: Detection and Inference. *Statistics in Medicine*, Vol.14, p.799–810.
- Loader, C. R., 1991, Large deviation approximations to the distribution of scan statistics. *Applied Probability Trust*, Vol.23, No.4, p.751-771.
- Naus, J. I. and Wallenstein, S., 2006, Temporal surveillance using scan statistics, Vol. 25, issue 2, p.311–324.
- Naus, J. I., 1974, Probabilities for generalized birthday problem. *Journal of the American Statistical Association*, Vol.69, p.810-815.
- Naus, J.I., 1965a, The distribution of the size of the maximum cluster of points on a line. *Journal of American Statistical Association*, Vol.60, p.532–538.
- Naus, J.I., 1965b, Clustering of random points in two dimensions, *Biometrika*, Vol.52, p.263-267.
- Öztürk, L., 2004, Monte-Carlo Simülasyon Metodu ve Bir İşletme Uygulaması, *Fırat Üniversitesi, DAUM Vol.3 No.1*, p:116-124.
- Sahu, S. K., Bendel, R. B. And Sison, C. P., 1993, Effect of relative risk and cluster configuration on the power of the one dimensional scan statistic, *Statistics in Medicine*, Vol.12, p.1853-1865.

- Saperstein, B., 1972, The generalized birthday problem. Journal of the American Statistical Association, Vol. 67, p. 425-428.
- Taş E., 2007, Simülasyon Modelleri, İstanbul Teknik Üniversitesi, http://web.itu.edu.tr/~yamanhak/ders/bdm/MM-0809-08_simulasyon.pdf.
- Tunbull B.W., Iwano E.J., Burnett W.S., Howe H.L. and Calrk L.C., 1990, Monitoring for clusters of disease: Application to leukemia incidence in upstate New York. Amerikan Journal of Epidemiology, p.132.
- Turner, R., Bottone, S. and Ghahramani Z., 29 Aug 2010 - 1 Sep 2010, Fast online anomaly detection using scan statistics, Machine Learning for Signal Processing (MLSP '10), Kittilä, Finland.
- Wallenstein, S., Naus, J. and Glaz, J., 1993, Power of scan statistic for detection of clustering, Statistics in Medicine, Vol.12, p.1829-1843.
- Wallenstein, S., Weinberg, C. R. and Gould, M., 1989, Testing for a pulse in seasonal event data, Biometrics, Vol.45, p.817-830.
- Walther, G., 2010, Optimal and fast detectin of spatial clusters with scan statistics, Annals of Statistics 2010, Vol.38, No.2, p.1010–1033.
- Weinstock, M.A., 1981, A generalized scan statistic test for the detection of clusters, International Journal of Edipemiology, Vol.10, p.289-293.

EKLER

EK 1: Tek Boyutlu Aylık Deprem Verileri Geriye Doğru Tarama

SUMMARY OF DATA

Study period.....: 2000/1 to 2011/12

Total number of cases.....: 82498

MOST LIKELY CLUSTER

Location IDs included.: All

Time frame.....: 2007/12 to 2011/12

Number of cases.....: 47103

Expected cases.....: 28082.82

Observed / expected...: 1.68

Relative risk.....: 2.58

Log likelihood ratio...: 9138.244798

P-value.....: 0.001

EK 2: Tek Boyutlu Aylık 4'ten Büyük Deprem Verileri Geriye Doğru Tarama

SUMMARY OF DATA

Study period.....: 2000/1/1 to 2011/12/31

Total number of cases.....: 1877

MOST LIKELY CLUSTER

Location IDs included.: All

Time frame.....: 2011/10 to 2011/11

Number of cases.....: 178

Expected cases.....: 26.12

Observed / expected...: 6.81

Relative risk.....: 7.42

Log likelihood ratio...: 196.108408

P-value.....: 0.001

EK 3: Tek Boyutlu Aylık Deprem Verileri İleriye Doğru Tarama

SUMMARY OF DATA

Study period.....: 2000/1 to 2011/12

Total number of cases.....: 82498

MOST LIKELY CLUSTER

Location IDs included.: All

Time frame.....: 2007/12 to 2011/12

Number of cases.....: 47103

Expected cases.....: 28082.82

Observed / expected...: 1.68

Relative risk.....: 2.58

Log likelihood ratio..: 9138.244798

P-value.....: 0.001

Recurrence interval...: 83 years

EK 4: Tek Boyutlu Aylık 4'ten Büyük Deprem Verileri İleriye Doğru Tarama

SUMMARY OF DATA

Study period.....: 2000/1 to 2011/12

Total number of cases.....: 1877

MOST LIKELY CLUSTER

Location IDs included.: All

Time frame.....: 2011/10 to 2011/12

Number of cases.....: 202

Expected cases.....: 39.40

Observed / expected...: 5.13

Relative risk.....: 5.62

Log likelihood ratio..: 174.991271

P-value.....: 0.001

Recurrence interval...: 83 years

EK 5: Tek Boyutlu 2009 Yılı Günlük Deprem Verileri Geriye Doğru Tarama

SUMMARY OF DATA

Study period.....: 2009/1/1 to 2009/12/31

Total number of cases.....: 15177

MOST LIKELY CLUSTER

Location IDs included.: All

Time frame.....: 2009/2/17 to 2009/3/4

Number of cases.....: 1276

Expected cases.....: 665.29

Observed / expected...: 1.92

Relative risk.....: 2.00

Log likelihood ratio...: 233.332535

P-value.....: 0.001

EK 6: Tek Boyutlu 2010 Yılı Günlük Deprem Verileri Geriye Doğru Tarama

SUMMARY OF DATA

Study period.....: 2010/1/1 to 2010/12/31

Total number of cases.....: 18697

MOST LIKELY CLUSTER

Location IDs included.: All

Time frame.....: 2010/11/1 to 2010/11/6

Number of cases.....: 573

Expected cases.....: 307.35

Observed / expected...: 1.86

Relative risk.....: 1.89

Log likelihood ratio...: 93.200723

P-value.....: 0.001

EK 7: Tek Boyutlu 2011 Yılı Günlük Deprem Verileri Geriye Doğru Tarama

SUMMARY OF DATA

Study period.....: 2011/1/1 to 2011/12/31

Total number of cases.....: 29577

MOST LIKELY CLUSTER

Location IDs included.: All
Time frame.....: 2011/10/23 to 2011/11/23
Number of cases.....: 6540
Expected cases.....: 2593.05
Observed / expected...: 2.52
Relative risk.....: 2.95
Log likelihood ratio..: 2407.080517
P-value.....: 0.001

EK 8: Tek Boyutlu 2009 Yılı Günlük Deprem Verileri İleriye Doğru Tarama

SUMMARY OF DATA

Study period.....: 2009/1/1 to 2009/12/31
Total number of cases.....: 15177

MOST LIKELY CLUSTER

Location IDs included.: All
Time frame.....: 2009/12/22 to 2009/12/31
Number of cases.....: 494
Expected cases.....: 415.81
Observed / expected...: 1.19
Relative risk.....: 1.19
Log likelihood ratio..: 7.137501
P-value.....: 0.004
Recurrence interval...: 250 days

EK 9: Tek Boyutlu 2010 Yılı Günlük Deprem Verileri İleriye Doğru Tarama

SUMMARY OF DATA

Study period.....: 2010/1/1 to 2010/12/31
Total number of cases.....: 18697

MOST LIKELY CLUSTER

Location IDs included.: All
Time frame.....: 2010/7/12 to 2010/12/31
Number of cases.....: 9103

Expected cases.....: 8861.87
Observed / expected...: 1.03
Relative risk.....: 1.05
Log likelihood ratio..: 6.231793
P-value.....: 0.007
Recurrence interval...: 143 days

EK 10: Tek Boyutlu 2011 Yılı Günlük Deprem Verileri İleriye Doğru Tarama

SUMMARY OF DATA

Study period.....: 2011/1/1 to 2011/12/31
Total number of cases.....: 29577

MOST LIKELY CLUSTER

Location IDs included.: All
Time frame.....: 2011/10/23 to 2011/12/31
Number of cases.....: 10436
Expected cases.....: 5672.30
Observed / expected...: 1.84
Relative risk.....: 2.30
Log likelihood ratio..: 2108.537186
P-value.....: 0.001
Recurrence interval...: 2.7 years

EK 11: İki Boyutlu 2009 Yılı Kümelenmeler

SUMMARY OF DATA

Study period.....: 2009/1/1 to 2009/12/31
Number of locations.....: 13852
Total population.....: 13865
Total number of cases.....: 47

MOST LIKELY CLUSTER

Coordinates / radius..: (36.460300 N, 27.042000 E) / 47.67 km
Population.....: 58
Number of cases.....: 6
Expected cases.....: 0.20

Observed / expected...: 30.52
Relative risk.....: 34.84
Log likelihood ratio..: 15.384846
P-value.....: 0.0044

EK 12: İki Boyutlu 2010 Yılı Kümelenmeler

SUMMARY OF DATA

Study period.....: 2010/1/1 to 2010/12/31
Number of locations.....: 17162
Total population.....: 17164
Total number of cases.....: 74

MOST LIKELY CLUSTER

Coordinates / radius..: (38.769000 N, 40.036300 E) / 6.23 km
Population.....: 215
Number of cases.....: 16
Expected cases.....: 0.93
Observed / expected...: 17.26
Relative risk.....: 21.75
Log likelihood ratio..: 32.726604
P-value.....: 0.0000000041

SECONDARY CLUSTERS

Coordinates / radius..: (36.000800 N, 27.242700 E) / 112.86 km
Population.....: 306
Number of cases.....: 11
Expected cases.....: 1.32
Observed / expected...: 8.34
Relative risk.....: 9.62
Log likelihood ratio..: 14.481885
P-value.....: 0.0072

EK 13: İki Boyutlu 2011 Yılı Kümelenmeler

SUMMARY OF DATA

Study period.....: 2011/1/1 to 2011/12/31

Number of locations.....: 28174

Total population.....: 28213

Total number of cases.....: 272

MOST LIKELY CLUSTER

Coordinates / radius..: (37.371000 N, 42.773700 E) / 189.78 km

Population.....: 7743

Number of cases.....: 187

Expected cases.....: 74.65

Observed / expected...: 2.51

Relative risk.....: 5.82

Log likelihood ratio..: 101.262054

P-value.....: < 0.000000000000000010

EK 14: Üç Boyutlu 2011 Yılı Kış Mevsimi Kümelenmeler

SUMMARY OF DATA

Study period.....: 2011/1/1 , 2011/2/28, 2011/12/31

Number of locations.....: 681

Total population.....: 681

Total number of cases.....: 32

MOST LIKELY CLUSTER

Coordinates / radius..: (38.881500 N, 42.566500 E) / 22.84 km

Time frame.....: 2011/2/22 to 2011/2/22

Population.....: 8

Number of cases.....: 5

Expected cases.....: 0.38

Observed / expected...: 13.30

Relative risk.....: 15.58

Log likelihood ratio..: 10.513511

P-value.....: 0.162

EK 15: Üç Boyutlu 2011 Yılı İlkbahar Mevsimi Kümelenmeler

SUMMARY OF DATA

Study period.....: 2011/3/1 to 2011/5/31

Number of locations.....: 913

Total population.....: 914

Total number of cases.....: 34

MOST LIKELY CLUSTER

Coordinates / radius..: (40.799300 N, 28.029200 E) / 206.30 km

Time frame.....: 2011/5/19 to 2011/5/19

Population.....: 14

Number of cases.....: 6

Expected cases.....: 0.52

Observed / expected...: 11.52

Relative risk.....: 13.78

Log likelihood ratio..: 10.983709

P-value.....: 0.084

EK 16: Üç Boyutlu 2011 Yılı Yaz Mevsimi Kümelenmeler

SUMMARY OF DATA

Study period.....: 2011/6/1 to 2011/8/31

Number of locations.....: 633

Total population.....: 633

Total number of cases.....: 22

MOST LIKELY CLUSTER

Coordinates / radius..: (39.120500 N, 29.043300 E) / 0.95 km

Time frame.....: 2011/6/27 to 2011/7/13

Population.....: 2

Number of cases.....: 2

Expected cases.....: 0.070

Observed / expected...: 28.77

Relative risk.....: 31.55

Log likelihood ratio..: 6.809489

P-value.....: 0.91

EK 17: Üç Boyutlu 2011 Yılı Sonbahar Mevsimi Kümelenmeler

SUMMARY OF DATA

Study period.....: 2011/9/1 to 2011/11/30

Number of locations.....: 2392

Total population.....: 2393

Total number of cases.....: 166

MOST LIKELY CLUSTER

Coordinates / radius.: (38.725000 N, 43.304200 E) / 22.11 km

Time frame.....: 2011/10/23 to 2011/10/26

Population.....: 427

Number of cases.....: 72

Expected cases.....: 29.62

Observed / expected...: 2.43

Relative risk.....: 3.53

Log likelihood ratio..: 31.800279

P-value.....: 0.000000046

ÖZGEÇMİŞ

Adı Soyadı : Rabia Burcu DEMİRTÜRK

Doğum Yeri : Çorlu/Tekirdağ

Doğum Yılı : 1987

Medeni Hali : Bekar

Eğitim ve Akademik Durumu:

Lise : 2001-2005 Samsun Atatürk Anadolu Lisesi

Lisans : 2005-2010 Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümü

Yabancı Dil: İngilizce