



T.C.
TOKAT GAZİOSMANPAŞA ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
MATEMATİK ANABİLİM DALI
YÜKSEK LİSANS PROGRAMI

BİNOMİK ORTALAMA

YÜKSEK LİSANS TEZİ
Mahmut ELMALI

Birinci Danışman: Prof. Dr. Naim ÇAĞMAN
İkinci Danışman: Dr. Dilek SABANCI

TOKAT- 2024

ETİK SÖZLEŞME

Tokat Gaziosmanpaşa Üniversitesi Lisansüstü Eğitim Enstitüsü tez yazım kılavuzuna göre, Prof. Dr. Naim ÇAĞMAN ve Dr. Dilek SABANCI danışmanlıklarında hazırlamış olduğum “Binomik Ortalama” adlı Yüksek Lisans tezinin bilimsel etik değerlere ve kurallara uygun, özgün bir çalışma olduğunu, aksinin tespit edilmesi halinde her türlü yasal yaptırımını kabul edeceğimi beyan ederim.

...../...../.....

Mahmut ELMALI

ÖNSÖZ

Yüksek lisans eğitim sürecimde ve yüksek lisans tezi olarak hazırladığım bu çalışmada, bilgi ve tecrübeleriyle her zaman katkı sunan ve desteklerini esirgemeyen tez danışmanlarım Sayın Prof. Dr. Naim ÇAĞMAN ve Sayın Dr. Dilek SABANCI hocalarıma çok teşekkür ederim.

Mahmut ELMALI

TOKAT-2024

ÖZET

BİNOMİK ORTALAMA

Elmalı, Mahmut

Yüksek Lisans, Matematik Anabilim Dalı

Tez Danışmanı: Prof. Dr. Naim Çağman

İkinci Danışman: Dr. Dilek Sabancı

Ağustos 2024, xi + 39 sayfa

Bu çalışmada, “binomik ortalama” adını verdiğimiz yeni bir ortalama önerdik. Binomik ortalama, en küçükten en büyüğe sıralanmış bir veri setindeki her bir gözlemin binom katsayısı ile çarpılması ile elde edilen ürünlerin toplamının binom katsayıları toplamına bölünmesiyle hesaplanır. Bu nedenle parametrik ortalamalarda olduğu gibi hesaplama, veri setindeki tüm gözlemler kullanılarak yapılır. Binom katsayıları uç değerlerden ortaya doğru büyüdüğünden, binomik ortalama, parametrik olmayan ortalamalarda olduğu gibi veri setindeki uç değerlerin etkisini azaltmıştır. Binomik ortalamanın istatistiksel ölçüm kalitesi açısından yansızlık, etkinlik, minimal yeterlilik, yeterlilik ve tutarlılık gibi özellikleri incelenmiştir. Farklı özelliklere sahip sürekli dağılımlardan 5000 örneklem oluşturulmuştur. Oluşturulan 5000 örneklemden her birinin medyan, binomik ortalama ve aritmetik ortalama sonuçlarının ortalaması ve varyansının karşılaştırılması yapılmıştır. Farklı veri kümelerine ait değişkenlerin hem olasılık dağılımlarının hem de medyan, binomik ortalama, aritmetik ortalama ve varyans tahminlerinin karşılaştırılması yapılmıştır. Elde edilen sonuçlar tablo ve grafiklerle verilmiştir. Düşük gözlemlerde diğer ortalamalardan daha düşük sonuçlar verirken, varyans tahmini açısından medyandan daha düşük sonuçlar vermiştir. Aşırı çarpık dağılımlarda dağılımın merkezine diğer ortalamalar göre daha iyi konumlanmıştır. Uçlardaki aykırı değerlerin varlığında medyandan daha etkin sonuçlar vermektedir. Bu sonuçlara göre binomik ortalama, istatistik alanında yapılacak çalışmalarda veri setlerinin yorumlanmasında özellikle parametrik olmayan ortalamalara alternatif bir ortalama olarak kullanılabilir.

Anahtar Kelimeler: Binomik Ortalama, Nokta Tahmini, Binom Dizisi, Binom Katsayısı, Ortalamalar

ABSTRACT

BİNOMİAL MEAN

Elmalı, Mahmut

Master's Thesis, Department of Mathematics

Advisor: Prof.Dr. Naim Çağman

Second Advisor: Dr. Dilek Sabancı

August 2024, xi + 39 pages

In this study, we propose a new mean, which we call the “binomial mean.” The binomial mean is calculated by multiplying each observation in a data set ordered from smallest to largest by a binomial coefficient and dividing the sum of the products by the sum of the binomial coefficients. Therefore, as with parametric means, the calculation is made using all observations in the data set. Since the binomial coefficients grow from the extremes towards the middle, the binomial mean reduces the effect of outliers in the data set, as in nonparametric means. The properties of binomial averaging in terms of statistical measurement quality such as unbiasedness, efficiency, minimal adequacy and consistency were analyzed. A comparison of the mean and variance of the median, binomial mean and arithmetic mean results of each of the 5000 samples generated from continuous distributions with different characteristics was made. Comparison of both probability distributions and median, binomial mean, arithmetic mean and variance estimates of variables belonging to different datasets were made. The results are given in tables and graphs. While it gives lower results than other averages for low observations, it gives lower results than the median in terms of variance estimation. In highly skewed distributions, it is better positioned at the center of the distribution than other means. It is more efficient than the median in the presence of extreme outliers. According to these results, the binomial mean can be used as an alternative mean to nonparametric means in the interpretation of data sets in studies to be conducted in the field of statistics.

Keywords: Binomial Mean, Point Estimation, Binomial Array, Binomial Coefficients, Means

İÇİNDEKİLER

ETİK SÖZLEŞME.....	i
ÖNSÖZ	v
ÖZET	vi
ABSTRACT.....	vii
İÇİNDEKİLER	vi
TABLolar LİSTESİ.....	viii
ŞEKİLLER LİSTESİ	ix
SİMGELER VE KISALTMALAR	x
1. GİRİŞ	1
2. TEMEL TANIM VE TEOREMLER.....	5
2.1. Aritmetik Ortalama	5
2.2. Geometrik Ortalama	5
2.3. Harmonik Ortalama	5
2.4. Kareli Ortalama.....	6
2.5. Medyan	6
2.6. Binom Katsayısı.....	7
3. BİNOMİK ORTALAMA	8
3.1. Binomik Ortalamanın Karşılaştırılması	9
3.1.1. Binomik Ortalamanın Aritmetik Ortalama ile Karşılaştırılması.....	9
3.1.2. Binomik Ortalama ile Ağırlıklı Ortalamanın Karşılaştırılması	11
4. BİNOMİK ORTALAMANIN TAHMİNCİ ÖZELLİKLERİNİN İNCELENMESİ..	12
4.1. Yansızlık	12
4.2. Etkinlik.....	14
4.3. Yeterlilik	18
4.4. Minimal Yeterlilik	20
4.5. Tutarlılık	21
5. UYGULAMALAR	23
5.1. Dağılım Simülasyonları	23
5.2. Veri Uygulamaları	25
SONUÇ.....	34

KAYNAKÇA.....	37
---------------	----



TABLÖLAR LİSTESİ

Tablo No	Sayfa
Tablo 1: Binom dizisinden binom ortalama formülünün elde edilmesi.....	8
Tablo 2: Farklı parametre değerlerine sahip normal dağılımların farklı gözlem değerlerinde 5000 örneklemin medyan, binom ortalama ve aritmetik ortalama sonuçlarının ortalama ve varyanslarının karşılaştırılması.....	26
Tablo 3: Farklı parametre değerlerine sahip gamma dağılımların farklı gözlem değerlerinde 5000 örneklemin medyan, binom ortalama ve aritmetik ortalama sonuçlarının ortalama ve varyanslarının karşılaştırılması.....	26
Tablo 4: Farklı parametre değerlerine sahip beta dağılımların farklı gözlem değerlerinde 5000 örneklemin medyan, binom ortalama ve aritmetik ortalama sonuçlarının ortalama ve varyanslarının karşılaştırılması.....	27
Tablo 5: Farklı veri setlerine ait değişkenlerin medyan, binomik ortalama, aritmetik ortalama ve varyans tahminlerinin karşılaştırılması.....	27

ŞEKİLLER LİSTESİ

Şekil No

Sayfa

Grafik 1: Farklı veri setlerine ait değişkenlerin olasılık yoğunluk fonksiyonları ve medyan, binom ortalaması ve aritmetik ortalama sonuçlarının karşılaştırılması....32



SİMGELER VE KISALTMALAR

Simgeler

Σ :	Toplam
$!$:	Faktöriyel
$\{ \}$:	Küme
$<$:	Küçüktür
$>$:	Büyüktür
μ :	Anakütle Ortalaması
$\bar{X}$:	Örneklem Ortalaması
B :	Anakütle Binomik Ortalaması
$\bar{B}$:	Örneklem Binomik Ortalaması
σ^2 :	Anakütle Varyansı
s^2 :	Örneklem Varyansı
$\Rightarrow$:	İse
$\int$:	İntegral
E :	Beklenen Değer
sd :	Standart Değişim

Kısaltmalar

Geo. O.	Geometrik Ortalama
Har. O.	Harmonik Ortalama
Kar. O.	Kareli Ortalama
M	Medyan

1. GİRİŞ

Kökleri antik Mezopotamya'ya (Ancient Mesopotamia) (MÖ 2000 civarı) kadar uzanan ortalama (mean) kavramının biçimsel gelişimi eski Yunanlılara atfedilse de matematiksel temeli ünlü matematikçi Pisagor'un teoremine dayanmaktadır. Öyle ki Pisagor ortalamaları (Pythagorean means) olarak bilinen aritmetik ortalama, geometrik ortalama ve harmonik ortalama Pisagorcular tarafından incelenmiştir (Heath, 1921). Ardından Yunan matematikçi Öklid, ünlü eseri "Elementler" de aritmetik ortalamayı tanımlamış ve özelliklerini tartışmıştır (Heath, 1956). Orta Çağ boyunca İslam alimleri Al-Khwarizmi ve Omar Hayyam, Rönesans'ta Leonardo Fibonacci gibi Avrupalı matematikçiler, 17. Yüzyılda Blaise Pascal ve John Wallis, 18. Yüzyılda ise Jacob Bernoulli ve Pierre-Simon Laplace gibi bilim adamları, bugün anladığımız şekliyle ortalama kavramına özel olarak odaklanmamışlardır. Ancak cebir ve geometriye olan katkıları, daha sonraki matematikçilerin ortalama ile ilgili kavramları keşfetmeleri ve geliştirmeleri için temel oluşturmuştur (Caillat, 1923; Scott, 1936; Struik, 1958; Dunlap Van der Waerden, 2013).

18. yüzyıl sonlarına doğru matematikten ayrılmaya başlayan istatistik biliminde meydana gelen gelişmeler ile matematikçiler ve istatistikçiler “merkezi eğilim” kavramını ileri sürmüşlerdir. İlgili dönemde merkezi eğilimin üç ana ölçüsü aritmetik ortalama, medyan ve mod olsa da geometrik ortalama, kareli ortalama ve harmonik ortalama gibi farklı ortalamalarda zamanla geliştirilmiştir. Bu kavramların pratik kullanımı uzun bir geçmişe sahip olsa da istatistiksel teori içindeki resmileştirilmesi ve tanınması zamanla gelişmiştir. Belçikalı istatistikçi Adolphe Quetelet, insan özelliklerine ilişkin çalışmalarında aritmetik ortalamayı kullanmış ve "ortalama insan" kavramını ortaya atarak merkezi eğilim ölçüsü olarak aritmetik ortalamanın kullanımını yaygınlaştırmıştır (Beirne, 1987). Bu çağın dikkate değer diğer isimleri Karl Pearson ve Francis Galton, büyük ölçüde aritmetik ortalamanın kullanımına dayanan en küçük kareler yöntemini geliştirmişlerdir (Hald, 1998). Medyan ölçüsüne aktüeryal uygulama için uygun olmayan bir istatistiğin örneği olarak Christiaan Huygens' in yazışmalarında rastlanmıştır (Bakker ve Gravemeijer,2006). Mod terimi ise 1895 yılında “maksimum frekansın ordinatına karşılık gelen apsisi” tanımlaması ile Karl Pearson tarafından ortaya atılmıştır (Pearson, 1895). 19. yüzyılda olasılık teorisi ve çıkarımsal istatistiğin

gelişmesiyle medyan ve mod daha fazla ilgi görmüştür ve resmi istatistiksel analizlerin bir parçası haline gelmiştir. Bu yüzyılın sonuna gelirken yine Karl Pearson, bir veri kümesindeki değişkenlik veya dağılım fikri ile ilgili olan kareli ortalama kavramını ortaya atmıştır. Dolayısıyla günümüzde oldukça sık kullandığımız varyans kavramının temeli bu dönemde oluşmuştur (Pearson,1936).

20. yüzyıl başlarına gelindiğinde çeşitli istatistikçilerin ve araştırmacıların katkılarıyla parametrik ve parametrik olmayan istatistik kavramları literatürde ki yerini almaya başlamıştır. Ronald Aylmer Fisher genellikle verilerin belirli bir dağılımdan geldiğini ve bu dağılımın parametrelerinin verilerden tahmin edildiğini varsayan, parametrik istatistik kavramının gelişimine önemli katkılarda bulunmuştur (Fisher, 1936). Arthur Lyon Bowley, Jacob Wolfowitz ve Samuel Stanley Wilks gibi isimler ise verilerin temel dağılımı hakkında spesifik varsayımlarda bulunmayan, parametrik olmayan istatistik kavramının gelişiminde önemli rol oynamışlardır (Sanger,1901; Mosteller, 1964; Zacks, 2003). Bu kavramların gelişimi ortalamaların, parametrik ve parametrik olmayan ortalamalar şeklinde ikiye ayrılmasına sebep olmuştur. Veri setindeki bütün değerlerin kullanıldığı ortalamalar parametrik, veri setindeki değerlerin sıralanıp bir veya birkaç değer kullanıldığı ortalamalar ise parametrik olmayan şeklinde ikiye ayrılmıştır (Conover, 1999; Pfanzagl, 2011; Sprent ve Smeeton, 2016). Dolayısıyla aritmetik, kareli, geometrik ve harmonik ortalama, parametrik ortalamalar grubuna girerken; mod, medyan ve kartiller ise parametrik olmayan ortalamalar grubunda yer almıştır.

Ortalama kavramı, 21. yüzyılın ikinci çeyreğine yaklaşırken modern istatistik, olasılık teorisi, programlama dilleri, yapay zekâ, makine öğrenimi vb. alanlardaki ilerlemelere bağlı olarak araştırmacılar tarafından geliştirilmeye devam etmektedir. Bilgi ve veri toplamanın önemi her geçen gün daha iyi anlaşılırken, toplanan bilgi ve verilerin doğru yorumlanmasına yönelik kullanılan teknikler daha da önem kazanmaktadır. Yorumlama sürecinde en çok kullanılan yöntemlerden birisi kuşkusuz ortalama hesaplamaktır. Aritmetik ortalama gibi parametrik bir ölçüm, veriler varsayılan dağılıma uygun olduğunda güçlüdür ancak aykırı değer varlığından etkilenmektedir. Medyan gibi parametrik olmayan bir ölçüm ise dağılım varsayımlarının belirsiz olduğu ya da ihlal edildiği durumlarda sağlamdır ve aykırı değerlerden daha az etkilenir. Bu iki belirgin farka sahip ortalamaların dezavantajlarını elemine eden ortalama ise çalışmamızın temelini oluşturan “Binomik Ortalama” kavramıdır.

Bu yeni ortalama temelinde, belli şartlar altında ağırlıklı ortalama hesaplamasında binom katsayılarının kullanılması esasına dayanmaktadır. Binom katsayılarının çeşitli varyasyonlarda kullanıldığı birçok çalışma, matematiksel ve istatistiksel bilginin işbirlikçi doğası nedeniyle literatürde bulunmaktadır. Son zamanlarda, Glasby ve Paseman binom toplamlarının maksimum dizisini ele almış ve bu dizinin tek modlu olduğunu ve hangi koşul altında maksimum değere ulaştığını göstermiştir (Glasby ve Paseman, 2021). Daha sonra Byun ve Poznanović, bu diziden doğal olarak ortaya çıkan ayrık olasılık dağılımları ailesini de dikkate alarak diziyi daha da genelleştirerek sonuçları incelemişlerdir (Byun ve Poznanović, 2024). Bu çalışmalarda binom katsayılarının kullanımı daha geniş bir kombinatorik ve matematiksel analiz bağlamında incelenmiştir. Kombinatorik, olasılık teorisi ve cebir ile ilgili alanlarda çalışan matematikçiler, binom katsayılarının anlaşılmasına ve uygulamalarına katkıda bulunmuşlardır (Wilson, 2016). Fakat binom katsayıları ile ağırlıklandırılmış bir ortalama olarak, hatta bir tahminleyici olarak istatistiksel açıdan diğer ortalamalar ile kıyaslanmış herhangi bir uygulamaya literatür de rastlanmamıştır. Çalışmamız ağırlıklı binomik toplamı bir diziden çıkarıp istatistiksel bir ortalama ve ölçüm formunda diğer ortalamalar ile ilişkisini göstermek açısından özgün bir değere sahiptir.

“Binomik ortalama” olarak isimlendirdiğimiz yeni ortalama, küçükten büyüğe doğru sıralanmış veri setinin her bir verisi, sırayla binom katsayıları ile çarpılır ve çarpımların toplamının binom katsayılarının toplamına bölümünden elde edilir. Binom katsayıları,

$$(x + y)^n = \binom{n}{0}x^n y^0 + \binom{n}{1}x^{n-1}y^1 + \binom{n}{2}x^{n-2}y^2 + \dots + \binom{n}{n}x^0 y^n$$

açılımından da görüldüğü gibi sıralamayı dikkate alırsak başlangıçtan ortaya doğru büyüyerek ortadan sonra küçülerek Pascal üçgenini oluşturmaktadır.

Binom katsayıları uç kısımlardan ortaya doğru büyüdüğü için binomik ortalama, veri setindeki aşırı uç değerlerin etkisini azaltmaktadır. Binomik ortalama, sapmasızlık, en küçük varyans ve etkinlik özelliklerini taşıyan, veri setindeki bütün değerleri kullandığı için parametrik ve veri setindeki değerleri sıraladığı için parametrik olmayan ortalama sayılabilir. Binomik ortalama, R Studio uygulama örnekleriyle aritmetik ortalama ve medyanla karşılaştırılarak, veri sayısı arttıkça aritmetik ortalama ile medyana göre daha az değişim gösterdiği ve aşırı uç değerlerden aritmetik ortalama göre daha az etkilendiği görülmüştür. Bu özellikleriyle binomik ortalama, istatistik alanında

yapılacak çalışmalarda veri setlerinin yorumlanması yapılırken parametrik ve parametrik olmayan ortalamalara alternatif bir ortalama olmaktadır.

Doğru yorumlama yapmayı engelleyen etkenler olabileceği gibi ortalama hesaplamayı engelleyen durumlarda vardır. Veri setini oluşturan değerler içerisinde sıfır bulunması halinde harmonik ve geometrik ortalama hesaplanmaz, negatif değer varsa harmonik ortalama hesaplanır ama anlamlı sonuç vermez ve geometrik ortalama hesaplaması yapılamaz olmasına rağmen binomik ortalama, bu tür bir engelleme yoktur.

Sonuç olarak bu çalışmanın temel konusunda, binomik ortalamanın genel tanımına, bir tahminleyici olarak istatistiksel özelliklerine (yansızlık, minimum varyans, etkinlik vb.), farklı gözlem değerlerine sahip dağılımlarda simülasyonlar yapılarak aritmetik ortalama ve medyana göre gösterdiği performansa yer verilmiştir.

2. TEMEL TANIM VE TEOREMLER

Bu bölümde aritmetik, geometrik, harmonik, kareli ortalamaların ve binom katsayısının tanımları verilecektir.

2.1. Aritmetik Ortalama

Bu bölümde aritmetik ortalama tanımı, özellikleri ve kullanışlı olmama sebepleri verilecektir.

Aritmetik ortalama en yaygın biçimde kullanılan ortalamadır. Bir dağılımda çok küçük ve çok büyük değerler olduğunda, bundan olumsuz etkilenmektedir. Bir serideki gözlem değerleri toplamının, toplam gözlem sayısına bölünmesiyle elde edilir.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

2.2. Geometrik Ortalama

Bu bölümde geometrik ortalama tanımı, özellikleri ve kullanışlı olmama sebepleri verilecektir.

Geometrik ortalama, bir seriyi oluşturan tüm değerlerin birbiriyle çarpılıp, değer sayısı kadar kökü kökünün alınmasıyla bulunan ortalamadır. Serideki aşırı değerlerden aritmetik ortalama kadar etkilenmediğinden, aşırı büyük veya aşırı küçük değerlere sahip serilerin ortalamasının bulunmasında kullanılmaktadır.

$$Geo.O. = \sqrt[n]{X_1 \cdot X_2 \dots X_n} = \sqrt[n]{\prod_{i=1}^n X_i}$$

2.3. Harmonik Ortalama

Bu bölümde harmonik ortalama tanımı, özellikleri ve kullanışlı olmama sebepleri verilecektir.

Harmonik ortalama, bir serideki gözlem değeri sayısının, serideki tüm değerlerin terslerinin toplamına bölünmesiyle bulunan ortalamadır. Oransal olarak verilen

değerlerin hesaplanmasında kullanılmaktadır. Bir seride sıfır varsa ve değerlerin işaretleri farklıysa harmonik ortalama kullanılmaz.

$$Har. O. = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}}$$

2.4. Kareli Ortalama

Bu bölümde kareli ortalama tanımı, özellikleri ve kullanışlı olmama sebepleri verilecektir.

Kareli ortalama, bir serideki tüm değerlerin karelerinin aritmetik ortalaması alınarak hesaplanan ve son aşamada karekök alma işlemiyle orijinal değerlere dönüşüm sağlandığı bir ortalama değildir. Kareli ortalama, aritmetik ortalama sapmaların bir ölçüsü olan standart sapmanın hesaplanmasında kullanıldığından ve negatif değerleri de dikkate aldığından, çok büyük bir öneme sahiptir.

$$Kar. O. = \sqrt{\frac{\sum_{i=1}^n X_i^2}{n}}$$

2.5. Medyan

Bu bölümde medyan tanımı ve özellikleri verilecektir.

Verileri küçükten büyüğe doğru sıraladıktan sonra ortadaki değerdir, n gözlem sayısı ve X veri kümesindeki değerlerin sıralı listesi olmak üzere, medyan değeri şu şekilde bulunur

$$M(X) = \begin{cases} X \left[\frac{n+1}{2} \right] & n \text{ tek sayı ise} \\ \frac{X \left[\frac{n}{2} \right] + X \left[\frac{n}{2} + 1 \right]}{2} & n \text{ çift sayı ise} \end{cases}$$

Bir olasılık dağılımı çarpıklık gösteriyorsa, medyan, aritmetik ortalama dan daha uygun bir merkezs el konum ölçüsüdür. Simetrik olmama durumu, sıralanmış veriler için ya en küçük değerlerin ya da en büyük değerlerin diğerlerinden çok daha fazla uzaklaşması durumunda ortaya çıkar. Diğer verilere göre küçük veya büyük değerlere aykırı

değer denir. Eğer veri dağılımı asitmetik olan aykırı değerler kapsıyorsa, medyan aritmetik ortalamaya nazaran daha güçlü bir merkezel konum ölçüsü durumuna gelir.

2.6. Binom Katsayısı

Bu bölümde binom katsayısının tanımı verilecektir.

Faktöriyel kavramının klasik tanımı $0! = 1$ olmak üzere $n! = n \cdot (n - 1) \cdot \dots \cdot 1$ şeklindedir. Faktöriyel tanımı kullanılarak, binom katsayıları da

$$\binom{n}{m} = \frac{n \cdot (n - 1) \cdot \dots \cdot 1}{m \cdot (m - 1) \cdot \dots \cdot 1 \cdot (n - m) \cdot (n - m - 1) \cdot \dots \cdot 1} = \frac{n!}{m! (n - m)!}$$

biçiminde tanımlanır.

Bu $\binom{n}{m}$ sayılarının “katsayı” olarak isimlendirilmesi, binom teorisine göre $(x + y)^n$ açılımında, $\sum_{r=0}^n \binom{n}{r} x^{n-r} y^r$ ifadesinde terimlerin katsayılarının binom sayıları olmasındandır.

3. BİNOMİK ORTALAMA

Bu bölümde yeni bir ortalama olan binomik ortalamanın tanımını ve temel özelliklerini vereceğiz.

Tanım 3.1. n pozitif bir tamsayı olmak üzere reel sayılar kümesinin bir alt kümesi olan $A = \{x_1, x_2, x_3, \dots, x_n\}$ sıralı kümesi bir veri seti olsun. A kümesinin elemanlarından elde edilen

$$\bar{B}(A) = \frac{1}{2^{n-1}} \sum_{i=1}^n \binom{n-1}{i-1} x_i \quad (3.1)$$

değere A veri setinin binomik ortalaması denir. Bundan sonra, binomik ortalamayı gerekli yerlerde $\bar{B}$ şeklinde kullanacağız. Burada, her $i=1,2,\dots,n$ için sırasıyla x_i verisinin $\binom{n-1}{i-1}$ binom katsayısı ile çarpılmasıyla, A veri kümesinin eleman sayısı 2^{n-1} olmaktadır.

Binom dizisinden binom ortalama formülünün elde edilmesi aşağıda verilen Tablo 1’de gösterilmiştir.

Tablo 1. Binom dizisinden binom ortalama formülünün elde edilmesi

N	Binom Üçgeni	Binomik Ortalama
1	$1 x_1$	$1 x_1$
2	$1 x_1 \quad 1 x_2$	$\frac{1}{2^1}(1 x_1 + 1 x_2)$
3	$1 x_1 \quad 2 x_2 \quad 1 x_3$	$\frac{1}{2^2}(1 x_1 + 2 x_2 + 1 x_3)$
4	$1 x_1 \quad 3 x_2 \quad 3 x_3 \quad 1 x_4$	$\frac{1}{2^3}(1 x_1 + 3 x_2 + 3 x_3 + 1 x_4)$
5	$1 x_1 \quad 4 x_2 \quad 6 x_3 \quad 4 x_4 \quad 1 x_5$	$\frac{1}{2^4}(1 x_1 + 4 x_2 + 6 x_3 + 4 x_4 + 1 x_5)$
$\vdots$	$\ddots \quad \ddots \quad \vdots \quad \ddots \quad \ddots$	$\vdots$
n	$1 x_1 \quad (n-1) x_2 \quad \dots \quad (n-1) x_{n-1} \quad 1 x_n$	$\frac{1}{2^{n-1}} \sum_{i=1}^n \binom{n-1}{i-1} x_i$

3.1. Binomik Ortalamanın Karşılaştırılması

Bu bölümde binomik ortalamanın aritmetik ve ağırlıklı aritmetik ortalama ile karşılaştırılması yapılacaktır.

3.1.1. Binomik Ortalamanın Aritmetik Ortalama ile Karşılaştırılması

Bu bölümde, binomik ortalama ile karşılaştırmasını yapmak için kullanılacak formül ve ispatı verilecektir.

Farklı ortalama yöntemlerinin sonuçlarını karşılaştırmak, uygunluk derecesi hakkında fikir oluşması açısından önemlidir. Bu çalışma boyunca aritmetik ortalamayı gerektiği yerlerde $\bar{X}$ şeklinde kullanacağız.

Örnek 3.1.1. $A = \{ a, a+k, a+2k, a+3k, a+4k, \dots, a+yk \}$ değerlerinden oluşan veri setinin binomik ortalama ve aritmetik ortalama değerlerini hesaplayalım.

Burada, binomik ortalama formülünü kullanırsak,

$$\begin{aligned}\bar{B}(A) &= \frac{2^y \cdot a + 2^{y-1} \cdot y \cdot k}{2^y} \\ &= a + \frac{y}{2} \cdot k\end{aligned}$$

Aritmetik ortalama formülünü kullanırsak,

$$\begin{aligned}\bar{X}(A) &= ((y+1) \cdot a + 2^{-1} \cdot y \cdot (y+1) \cdot k) / (y+1) \\ &= a + \frac{y}{2} \cdot k\end{aligned}$$

Burada, sabit artışlı veri setlerinde aritmetik ve binom ortalamalarının eşit olduğu görülmektedir.

Örnek 3.1.2. $B = \{ a, a+k, a+2k, a+3k, a+14k \}$ değerlerinden oluşan veri setinin binomik ortalama ve aritmetik ortalama değerlerini hesaplayalım.

Burada, binomik ortalama formülünü kullanırsak,

$$\bar{B}(B) = \frac{\binom{4}{0} \times a + \binom{4}{1} \times (a + k) + \binom{4}{2} \times (a + 2k) + \binom{4}{3} \times (a + 3k) + \binom{4}{4} \times (a + 14k)}{2^4}$$

$$= (16a+42k)/16 = a+2.625k$$

Aritmetik ortalama formülünü kullanırsak,

$$\bar{X}(B) = (a+a+k+a+2k+a+3k+a+14k)/5 = a+4k$$

Burada, aşırı aykırı değerlerin bulunduğu veri setinde, aritmetik ortalamanın binomik ortalamaya göre aşırı aykırı değerlerden daha fazla etkilendiği görülmektedir.

Örnek 3.1.3. $C = \{1, 34, 67, 100, 133, 166\}$ sayılarından oluşan veri setinin binomik ortalama ve aritmetik ortalama değerlerini hesaplayalım. Burada,

$$\bar{B}(C) = \frac{\binom{5}{0} \times 1 + \binom{5}{1} \times 34 + \binom{5}{2} \times 67 + \binom{5}{3} \times 100 + \binom{5}{4} \times 133 + \binom{5}{5} \times 166}{2^5}$$

$$= \frac{1 \times 1 + 5 \times 34 + 10 \times 67 + 10 \times 100 + 5 \times 133 + 1 \times 166}{32} = 83.5$$

$$\bar{X}(C) = \frac{1+34+67+100+133+166}{6} = 83.5$$

Görülmektedir ki artış miktarı eşit olunca binomik ortalama ile aritmetik ortalama eşit çıkmıştır.

Örnek 3.1.4. $D = \{1, 2, 3, 4, 5, 100\}$ sayılarından oluşan veri setinin binomik ortalama ve aritmetik ortalama değerlerini hesaplayalım. Burada,

$$\bar{B}(D) = \frac{\binom{5}{0} \times 1 + \binom{5}{1} \times 2 + \binom{5}{2} \times 3 + \binom{5}{3} \times 4 + \binom{5}{4} \times 5 + \binom{5}{5} \times 100}{2^5}$$

$$= \frac{1 \times 1 + 5 \times 2 + 10 \times 3 + 10 \times 4 + 5 \times 5 + 1 \times 100}{32} = 6.4$$

$$\bar{X}(D) = \frac{1+2+3+4+5+100}{6} = 19.2$$

Görülmektedir ki veri setinde aykırı değer varsa ve değerler sağa çarpık bir dağılım oluşturuyorsa binomik ortalama, aritmetik ortalamadan küçük çıkmıştır ve binomik ortalama verilerin yığıldığı kısma yakın bir değer almıştır.

Örnek 3.1.5. $E = \{1, 96, 97, 98, 99, 100\}$ sayılarından oluşan veri setinin binomik ortalama ve aritmetik ortalama değerlerini hesaplayalım. Burada,

$$\begin{aligned}\bar{B}(E) &= \frac{\binom{5}{0} \times 1 + \binom{5}{1} \times 96 + \binom{5}{2} \times 97 + \binom{5}{3} \times 98 + \binom{5}{4} \times 99 + \binom{5}{5} \times 100}{2^5} \\ &= \frac{1 \times 1 + 5 \times 96 + 10 \times 97 + 10 \times 98 + 5 \times 99 + 1 \times 100}{32} = 94.6\end{aligned}$$

$$\bar{X}(E) = \frac{1+96+97+98+99+100}{6} = 81.8$$

Görülmektedir ki veri setinde aykırı değer varsa ve değerler sola çarpık bir dağılım oluşturuyorsa binomik ortalama, aritmetik ortalamadan büyük çıkmıştır ve binomik ortalama verilerin yığıldığı kısma yakın bir değer almıştır.

Örnek 3.1.1, 3.1.2, 3.1.3, 3.1.4 ve 3.1.5 incelendiğinde binomik ortalamanın değerlerin yoğun olduğu sayılara aritmetik ortalamaya göre daha yakın sonuçlar verdiği görülmüştür. Aritmetik ortalama ise binomik ortalamaya göre aşırı uç değerlerin olduğu yere daha yakın sonuçlar vermiştir.

3.1.2. Binomik Ortalama ile Ağırlıklı Ortalamanın Karşılaştırılması

Bu bölümde, binomik ortalama ile ağırlıklı ortalamanın karşılaştırılması yapılacaktır.

Veri setini oluşturan değerler eşit öneme sahip değilse aritmetik ortalama yerine ağırlıklı ortalama kullanılır.

$X = \{x_1, x_2, x_3, \dots, x_n\}$ veri seti olsun. x_i verisinin katsayısı w_i ise X veri setinin ağırlıklı ortalaması,

$$\bar{X}_w = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}$$

şeklinde tanımlıdır.

Ağırlıklı ortalama hesaplanırken verilerin küçükten büyüğe doğru sıralanmasına gerek yoktur. Her örnek durumun ağırlıklı ortalaması hesaplanırken farklı katsayılar kullanılır ama binomik ortalama hesaplanırken katsayılar sabit kalır.

4. BİNOMİK ORTALAMANIN TAHMİNCİ ÖZELLİKLERİNİN İNCELENMESİ

Bu bölümde binomik ortalamanın; aritmetik ortalama, medyan ve varyans vb. tahmin edicilerde bulunması gereken özellikleri taşıdığı araştırılacaktır.

Bir parametre için değişik tahminler önerilebileceği gibi bir tahminden daha iyi tahminler de önerilebilir. Bu nedenle iyi bir tahminleyicinin bazı ölçüm kriterlerini sağlaması gereklidir. Ancak sağlanan bu kriterler kapsamında iyi bir tahminden söz edilebilir. Bu bölümde, binomik ortalamanın bir tahminleyici olarak hangi özelliklere sahip olduğu incelenecektir. Kesikli ve sürekli birçok dağılımın normal dağılıma yaklaşımının olduğu düşüncesi ile örnekler normal dağılım üzerinden verilmiştir.

Bu bölümde verilen yansızlık, etkinlik, yeterlilik, minimal yeterlilik ve tutarlılık tanımları (Akdi, 2011 ; Hogg, McKean ve Craig, 2013) kaynakları dikkate alınarak yazılmıştır.

4.1. Yansızlık

Bütün verilerin oluşturduğu kümeye anakütle, anakütle içerisinde kullanılmak üzere seçilen her alt kümeye örneklem denir. Anakütle A ve örneklem de X ve indisleri ile gösterilir.

Her örneklem aritmetik ortalama, varyans, oran vb. değerlerini veren fonksiyona tahmin edici veya istatistik denir. Her tahmin edicinin aldığı değere bir tahmin denir.

Tanım 4.1.1. X rasgele değişkeni $x_1, x_2, x_3, \dots, x_n, \dots$ mümkün değerlerini $p(x_i) = P(X = x_i), i = 1, 2, \dots, n, \dots$ olasılıklarıyla alan kesikli bir rasgele değişken olsun. Bu takdirde X rasgele değişkeninin $E(X)$ ile gösterilen beklenen değeri (veya matematiksel beklentisi)

$$E(X) = \sum_{i=1}^{\infty} x_i \cdot p(x_i)$$

olarak tanımlanır. Burada $\sum_{i=1}^{\infty} x_i \cdot p(x_i)$ serisi mutlak yakınsak yani $\sum_{i=1}^{\infty} |x_i| \cdot p(x_i) < \infty$ olmalıdır. Bu sayıya X 'in ortalama değeri olarak da müracaat edilir.

Tanım 4.1.2. X rastgele değişkeni f olasılık yoğunluk fonksiyonuna sahip bir sürekli rastgele değişken olsun. Bu durumda X rastgele değişkeninin beklenen değeri

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

olarak tanımlanır. Yine bu genelleştirilmiş integral yakınsak olmayabilir. Bu nedenle $E(X)$ 'in mevcut olması için gerek yeter koşul

$$\int_{-\infty}^{+\infty} |x| f(x) dx$$

integralinin sonlu olmasıdır.

Tanım 4.1.3. T tahmin edicisinin örneklem dağılımı beklenen değerinin (ortalamasının), ana kütle parametresi θ 'ya eşit olmasıdır. Yani,

$$E(T) = \theta$$

ise, T , θ 'nın sapmasız bir tahmincisidir. Buna göre, bir tahmin edicinin sapması, beklenen değeri ile tahmin edilmek istenen parametre θ arasındaki fark olarak,

$$\text{Sapma}(T) = E(T) - \theta$$

şeklinde gösterilir.

Eğer tahmin edici aritmetik ortalama ise " $E(\bar{X}) = \mu$ "dür.

Sapmasız bir tahmin edici için sapma miktarı sıfıra eşit olur. Parametrenin ortalama olarak gerçek değerini elde edebilmek için kullanılacak en iyi tahmin edici, sapmasız bir tahmin edicidir.

Beklenen değer; eldeki verilere göre kriterlerin olması muhtemel en olası değerini ifade etmektedir (Karacasu ve Yayla, 2004).

Teorem 4.1.1 $X_1, X_2, \dots, X_n$ beklenen değeri μ ve varyansı σ^2 olan normal dağılımdan bir örneklem olmak üzere $T = \frac{1}{2^{n-1}} \sum_{i=1}^n \binom{n-1}{i-1} x_i$ tahmin edicisi, μ 'nün yansız bir tahminleyicisidir.

İspat:

$$\begin{aligned}
 E(T) &= E \left[\frac{1}{2^{n-1}} \sum_{i=1}^{n-1} \binom{n-1}{i-1} x_i \right] \\
 &= \frac{1}{2^{n-1}} E \left[\sum_{i=1}^{n-1} \binom{n-1}{i-1} x_i \right] \\
 &= \frac{1}{2^{n-1}} E \left[\binom{n-1}{0} x_1 + \binom{n-1}{1} x_2 + \dots + \binom{n-1}{n-1} x_n \right] \\
 &= \frac{1}{2^{n-1}} \left[\binom{n-1}{0} E(x_1) + \binom{n-1}{1} E(x_2) + \dots + \binom{n-1}{n-1} E(x_n) \right] \\
 &= \frac{1}{2^{n-1}} \left[\binom{n-1}{0} \mu + \binom{n-1}{1} \mu + \dots + \binom{n-1}{n-1} \mu \right] \\
 &= \frac{1}{2^{n-1}} \cdot \mu \cdot \left[\left(\binom{n-1}{0} + \binom{n-1}{1} + \dots + \binom{n-1}{n-1} \right) \right] \\
 &= \frac{1}{2^{n-1}} \cdot \mu \cdot 2^{n-1} \\
 &= \mu
 \end{aligned}$$

4.2. Etkinlik

Varyans, verilerin aritmetik ortalamadan farklarının karelerinin toplamının veri sayısına oranıdır. En küçük varyans ölçütü, tahmin ediciler arasından tahmin hatasının varyansını en küçük yapan tahmin edicinin tercih edilmesi gerektiğini ifade eder. Örneklem dağılımın varyansı daha küçük olan tahmin edici daha iyi bir tahmin edicidir. Varsa en etkin tahmin edici kullanılmak istenir.

σ^2 anakütle varyansı, N anakütle veri sayısı ve s^2 örneklem varyansı, n örneklem veri sayısı olmak üzere anakütle ve örneklem varyansları,

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N} \quad , \quad s^2 = \frac{\sum (x - \bar{X})^2}{n - 1}$$

şeklindedir.

Tanım 4.2.1. T_1 ve T_2 , herhangi bir θ parametresi için iki tahmin edici olsun. Her θ için

$$Var_{\theta}(T_1) \leq Var_{\theta}(T_2)$$

oluyorsa T_1 tahmin edicisi T_2 tahmin edicisine göre daha etkindir denir.

Teorem 4.2.1 $X_1, X_2, \dots, X_n$ normal dağılıma sahip $X \sim N(\mu, \sigma^2)$ bir örneklem olsun. Buna göre, örneklem ortalaması ($\bar{X}$), örneklem binomik ortalaması $\bar{B}$ ve verilerin toplam değeri için,

$$Var(\bar{X}) \leq Var(\bar{B}) \leq Var\left(\sum x_i\right)$$

eşitsizliği elde edilir.

İspat:

$$\begin{aligned} E(T_1) &= E(\bar{X}) = E\left[\frac{\sum_{i=1}^n x_i}{n}\right] = E\left[\frac{x_1 + x_2 + \dots + x_n}{n}\right] \\ &= \frac{1}{n} [E(x_1) + E(x_1) + \dots + E(x_n)] \\ &= \frac{1}{n} [\mu + \mu + \dots + \mu] \\ &= \frac{1}{n} \cdot n \cdot \mu \\ &= \mu \end{aligned}$$

$$\begin{aligned}
Var(T_1) &= Var(\bar{X}) \\
&= Var\left(\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^n x_i\right) \\
&= \frac{1}{n^2} Var(x_1 + x_2 + \dots + x_n) \\
&= \frac{1}{n^2} (Var(x_1) + Var(x_2) + \dots + Var(x_n)) \\
&= \frac{1}{n^2} (\sigma^2 + \sigma^2 + \dots + \sigma^2) \\
&= \frac{1}{n^2} \cdot n \cdot \sigma^2 \\
&= \frac{\sigma^2}{n} \tag{4.2.1}
\end{aligned}$$

olarak elde edilir. O halde $T_1 = \bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} \sim N(\mu, \frac{\sigma^2}{n})$ elde edilir.
 $T_2 = \sum_{i=1}^n x_i$ tahmin edicisinin beklenen değeri ve varyansı;

$$\begin{aligned}
E\left(\sum_{i=1}^n x_i\right) &= E[x_1 + x_2 + \dots + x_n] = E(x_1) + E(x_2) + \dots + E(x_n) = \mu + \mu + \dots + \mu \\
&= n\mu
\end{aligned}$$

$$\begin{aligned}
Var\left(\sum_{i=1}^n x_i\right) &= Var[x_1 + x_2 + \dots + x_n] = Var(x_1) + Var(x_2) + \dots + Var(x_n) \\
&= \sigma^2 + \sigma^2 + \dots + \sigma^2 = n\sigma^2
\end{aligned}$$

olarak elde edilir. O halde, $T_2 = \sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n \sim N(n\mu, n\sigma^2)$ elde edilir.

$T_3 = \frac{1}{2^{n-1}} \sum_{i=1}^n \binom{n-1}{i-1} x_i$ tahmin edicisinin beklenen değeri ve varyansı; $E(\bar{B}) = \mu$ olarak

yansızlık örneğinde elde edildi.

$$\begin{aligned}
Var(\bar{B}) &= Var\left(\frac{1}{2^{n-1}} \sum_{i=1}^n \binom{n-1}{i-1} x_i\right) = \frac{1}{2^{2(n-1)}} Var\left(\sum_{i=1}^n \binom{n-1}{i-1} x_i\right) \\
&= \frac{1}{2^{2(n-1)}} Var\left(\binom{n-1}{0} x_1 + \binom{n-1}{1} x_2 + \dots + \binom{n-1}{n-1} x_n\right) \\
&= \frac{1}{2^{2(n-1)}} \left(Var\left(\binom{n-1}{0} x_1\right) + Var\left(\binom{n-1}{1} x_2\right) + \dots \right. \\
&\quad \left. + Var\left(\binom{n-1}{n-1} x_n\right)\right) \\
&= \frac{1}{2^{2(n-1)}} \left[\binom{n-1}{0}^2 Var(x_1) + \binom{n-1}{1}^2 Var(x_2) + \dots \right. \\
&\quad \left. + \binom{n-1}{n-1}^2 Var(x_n)\right] \\
&= \frac{1}{2^{2(n-1)}} \left[\binom{n-1}{0}^2 \sigma^2 + \binom{n-1}{1}^2 \sigma^2 + \dots + \binom{n-1}{n-1}^2 \sigma^2\right] \\
&= \frac{\sigma^2}{2^{2(n-1)}} \left[\binom{n-1}{0}^2 + \binom{n-1}{1}^2 + \dots + \binom{n-1}{n-1}^2\right] \\
&= \frac{\sigma^2}{2^{2(n-1)}} \left[\sum_{i=0}^n \binom{n-1}{i}^2\right] = \frac{\sigma^2}{2^{2(n-1)}} \binom{2n-2}{n-1} \tag{4.2.2}
\end{aligned}$$

olarak elde edilir. O halde $T_3 = \bar{B} = \frac{1}{2^{n-1}} \sum_{i=1}^n \binom{n-1}{i-1} x_i \sim N\left(\mu, \frac{\sigma^2}{2^{2n-2}} \binom{2n-2}{n-1}\right)$ elde edilir.

Birbirlerinden farklı bu tahmin edicilerin aralarındaki ilişki şu şekildedir;

$$Var(\bar{X}) \leq Var(\bar{B}) \leq Var\left(\sum x_i\right)$$

$$Var(T_1) \leq Var(T_3) \leq Var(T_2)$$

elde edilir. Bu sonuca göre $T_3 = \bar{B}$ tahmin edicisi $T_2 = \sum x_i$ tahmin edicisinden daha etkin iken, $T_1 = \bar{X}$ tahmin edicisinden etkin değildir.

Örnek 4.2.1. Örnek 3.1, 3.2 ve 3.3'teki veri setlerinin varyanslarını karşılaştıralım.

Örnek 3.1'deki veri setini kullanırsak;

$A = \{1, 34, 67, 100, 133, 166\}$ $\bar{B}(A) = 83.5$ $\bar{X}(A) = 83.5$ ise
 $Var(\bar{X}) = Var(\bar{B})$ olacaktır.

Yapılan varyans hesaplamaları sonucunda,

$Var(\bar{X}(A)) = 3176.3$ ve $Var(\bar{B}(A)) = 3176.3$ elde edilir ve
 $Var(\bar{X}) = Var(\bar{B})$ olur.

Örnek 3.2'deki veri setini kullanırsak;

$B = \{1, 2, 3, 4, 5, 100\}$ $\bar{B}(A) = 6.4$ $\bar{X}(A) = 19.2$ ise
 $Var(\bar{X}) < Var(\bar{B})$ olacaktır.

Yapılan varyans hesaplamaları sonucunda,

$Var(\bar{X}(B)) = 1308.5$ ve $Var(\bar{B}(B)) = 1470.5$ elde edilir ve
 $Var(\bar{X}) < Var(\bar{B})$ olur.

Örnek 3.3'teki veri setini kullanırsak;

$C = \{1, 96, 97, 98, 99, 100\}$ $\bar{B}(C) = 94.6$ $\bar{X}(C) = 81.8$ ise
 $Var(\bar{X}) < Var(\bar{B})$ olacaktır.

Yapılan varyans hesaplamaları sonucunda

$Var(\bar{X}(C)) = 1308.5$ ve $Var(\bar{B}(C)) = 1470.5$ elde edilir ve
 $Var(\bar{X}) < Var(\bar{B})$ olur.

4.3. Yeterlilik

Örnekleme için seçilen veriler, parametre hakkında ne kadar bilgi varsa, bilgilerin tamamını özetleyen bir tahmin edici, o parametre için yeterlidir.

$X_1, X_2, \dots, X_n$ olasılık veya olasılık yoğunluk fonksiyonu $f(x; \theta)$ olan kitleden bir örneklem olsun.

Tanım 4.3.1 $T(X)$ verildiğinde, X in koşullu dağılımı θ parametresine bağlı değil ise, $T(X)$ e θ için yeterli bir tahmin edici denir (Akdi, 2011 ; Hogg, McKean ve Craig, 2013).

Teorem 4.3.1 T tahmin edicisinin θ için yeterli olması için gerek ve yeter koşul $p(x; \theta)/q(T(x); \theta)$ oranının θ dan bağımsız olmasıdır.

Teorem 4.3.2 $X_1, X_2, \dots, X_n$ beklenen değeri μ ve varyansı 1 olan normal dağılımdan bir örnek olsun. O halde $X_i \sim N(\mu, 1)$ olsun. Teorem 4.2.1'den $T^* = X_1, X_2, X_3, \dots, X_n \sim N(n\mu, 1)$ 'dır. μ için T^* yeterlidir.

İspat: Teorem 4.3.1'deki oranı kullanırsak,

Pay,

$$\begin{aligned} p(\underline{X}; \mu) &= \prod_{i=1}^n f_{\mu}(X_i; \mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(X_i - \mu)^2\right) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n (X_i - \mu)^2\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n X_i^2 + \mu \sum_{i=1}^n X_i - \frac{n\mu^2}{2}\right) \end{aligned}$$

ve payda,

$$\begin{aligned} q(T(\underline{X}); \mu) &= q(t; \mu) = \frac{1}{\sqrt{2\pi n}} \exp\left(-\frac{1}{2n}(t - n\mu)^2\right) \\ &= \frac{1}{\sqrt{2\pi n}} \exp\left(-\frac{1}{2n}(t^2 - 2n\mu t + n^2\mu^2)\right) \\ &= \frac{1}{\sqrt{2\pi n}} \exp\left(-\frac{t^2}{2n} - \frac{n\mu^2}{2n} + \mu t\right) \end{aligned}$$

böylece,

$$\begin{aligned} \frac{p(\underline{X}; \mu)}{q(T(\underline{X}); \mu)} &= \frac{p(\underline{X}; \mu)}{q(t; \mu)} \\ &= \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n (X_i^2 + \mu \sum_{i=1}^n X_i - n\mu^2/2)\right)}{\frac{1}{\sqrt{2\pi n}} \exp\left(-\frac{t^2}{2n} + \mu t - n\mu^2/2\right)} \\ &= \frac{\sqrt{2n\pi}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \left[\sum_{i=1}^n X_i^2 - \frac{t^2}{n} \right]\right) \end{aligned}$$

bu oran μ 'ye bağlı olmadığından, T^* , μ için yeterli bir istatistiktir.

4.4. Minimal Yeterlilik

Bir parametre için sadece bir tane yeterli istatistik olmadığını, aynı zamanda yeterli istatistiklerin her bire bir fonksiyonunun aynı parametre için yeterli olduğunu biliyoruz. Bu nedenle, bir parametre için birçok yeterli istatistik önerilebilir.

Tanım 4.4.1 Bir T^* istatistiği, herhangi bir diğer T yeterli istatistiği T^* 'nin bir fonksiyonu olarak ifade edilebiliyorsa, bir θ parametresi için minimum yeterli olarak adlandırılır. Başka bir deyişle, T^* diğer herhangi bir yeterli istatistik T 'nin bir fonksiyonudur (Akdi, 2011 ; Hogg, McKean ve Craig, 2013).

Teorem 4.4.1 $X_1, X_2, \dots, X_n$ olasılık yoğunluk fonksiyonu $f(X_i, \theta)$ olan bir popülasyondan alınan bir örnek olsun. Eğer

$$\frac{f(\underline{X}; \theta)}{f(\underline{Y}; \theta)} = g(\underline{X}; \underline{Y})$$

'den bağımsızdır ve $T(\underline{X}) = T(\underline{Y})$ ise T, θ için minimal bir yeterli istatistiktir.

Teorem 4.3.2'nin ispatında $T^* = \sum_{i=1}^n X_i$ 'nin μ için yeterli olduğunu göstermiştik. Çünkü $T = \frac{1}{2^{n-1}} \sum_{i=1}^n \binom{n-1}{i-1} X_i$, T^* 'nin bir fonksiyonudur, T 'nin μ için minimum yeterli bir istatistik olduğu sonucuna varılır.

İspat: $X_1, X_2, \dots, X_n$, beklentisi μ ve varyansı 1 olan $N(\mu, 1)$ normal dağılımından bir örnek olsun. Şunu gözlemleyebiliriz:

$$\begin{aligned} \frac{f(\underline{X}; \mu, 1)}{f(\underline{Y}; \mu, 1)} &= \frac{\left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n X_i^2 + \mu \sum_{i=1}^n X_i - n\mu^2/2\right)}{\left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n Y_i^2 + \mu \sum_{i=1}^n Y_i - n\mu^2/2\right)} \\ &= \frac{\exp\left(-\frac{1}{2} \sum_{i=1}^n X_i^2 + \mu \sum_{i=1}^n X_i\right)}{\exp\left(-\frac{1}{2} \sum_{i=1}^n Y_i^2 + \mu \sum_{i=1}^n Y_i\right)} \\ &= \exp\left(-\frac{1}{2} \left[\sum_{i=1}^n X_i^2 - \sum_{i=1}^n Y_i^2 \right] + \mu \left[\sum_{i=1}^n X_i - \sum_{i=1}^n Y_i \right]\right) \end{aligned}$$

bu oranın μ parametresinden bağımsız olması için gerekli ve yeterli koşul şudur:

$$\sum_{i=1}^n X_i = \sum_{i=1}^n Y_i$$

Yani, $\underline{T}(\underline{X}) = (\sum_{i=1}^n X_i, 1)$ için gerekli ve yeterli koşul $\frac{f(\underline{X}; \mu, 1)}{f(\underline{Y}; \mu, 1)}$ oranının μ parametresinden bağımsız olması için gerekli koşul $\underline{T}(\underline{X}) = \underline{T}(\underline{Y})$ 'dir. Teorem 4.4.1'e göre, $\underline{T}(\underline{X}) = (\sum_{i=1}^n X_i, 1)$ istatistiği $(\mu, 1)$ için minimum yeterlidir.

İspattan elde edilen sonuca göre,

$$\underline{T}^*(\underline{X}) = \left(\frac{1}{2^{n-1}} \sum_{i=1}^n \binom{n-1}{i-1} X_i, 1 \right)$$

aynı zamanda $(\mu, 1)$ için minimum yeterlidir.

4.5. Tutarlılık

Bir tahmin edici örneklemin herhangi bir fonksiyonudur. Dolayısı ile, tahmin edici örneklem hacmine de bağlıdır (Akdi, 2011 ; Hogg, McKean ve Craig, 2013).

Tanım 4.5.1 $X_1, X_2, \dots, X_n$ olasılık veya olasılık yoğunluk fonksiyonu $f(x; \theta)$ olan kitleden bir örneklem, $T_n = T_n(\underline{X})$ de θ nın herhangi bir tahmin edicisi olsun. Eğer, $n \rightarrow \infty$ iken $T_n \xrightarrow{P} \theta$ ise, T_n tahmin edicilerinin dizisi θ için tutarlıdır denir.

Teorem 4.5.1 T_n bir tahmin ediciler dizisi olsun. Eğer

$$\lim_{n \rightarrow \infty} Var_{\theta}(T_n) = 0 \text{ ve } \lim_{n \rightarrow \infty} [E_{\theta}(T_n) - \theta]^2 = 0 ,$$

sağlandığında, T_n tahmin edicileri dizisi θ için tutarlıdır.

İspat: Ortalaması μ ve varyansı σ^2 olan bir normal dağılımdan alınan $X_1, X_2, \dots, X_n$ örnekleri için istatistiğin

$$T_n = \frac{1}{2^{n-1}} \sum_{i=1}^{n-1} X_i$$

μ için minimum düzeyde yeterlidir.

Teorem 4.1.1 ve teorem 4.2.1'de sırayla bu istatistiğin beklenen değerini ve varyanslarını tartışmıştık. Bu bilgilere dayanarak, T_n tahmin ediciler dizisi için:

$$\lim_{n \rightarrow \infty} Var_{\mu}(T_n) = \lim_{n \rightarrow \infty} \left(\frac{\sigma^2}{2^{2n-2}} \binom{2n-2}{n-1} \right) = 0$$

$$\lim_{n \rightarrow \infty} [E_{\mu}(T_n) - \mu]^2 = \lim_{n \rightarrow \infty} (\mu - \mu)^2 = 0$$

Dolayısıyla, T_n tahminci dizisi μ için tutarlıdır.

Binomik ortalama aritmetik ortalamayı belli oranda içermektedir. Buna bağlı olarak aritmetik ortalamanın sağladığı kalite ölçütlerini, binomik ortalama da sağlamaktadır.



5. UYGULAMALAR

5.1. Dağılım Simülasyonları

Bu bölümde Normal, Gamma ve Beta gibi farklı istatistiksel özelliklere sahip dağılımlar kullanılarak R Studio (version 4.4.1) programında simülasyon çalışması gerçekleştirilmiştir. Bu dağılımlardan herhangi birinin belli parametre değerleri ve gözlem sayısı için 5000 tekrar yapılarak örneklemeler oluşturulmuştur. Oluşturulan her bir örneklemin Medyan (M), Binomik Ortalama ($\bar{B}$) ve Aritmetik Ortalama ($\bar{X}$) değerleri hesaplanmıştır. Dolayısıyla elimizde 5000 adet M , 5000 adet $\bar{B}$ ve 5000 adet $\bar{X}$ sonucundan elde edilmiş üç farklı veri seti oluşmuştur. Bu veri setlerinin de ortalama ve varyans değerleri hesaplanarak sonuçlar elde edilmiştir. Sürekli dağılım ailesine sahip bu dağılımları seçmemizin birinci sebebi, farklı kesikli dağılımların bu dağılımlara yaklaşımlarının olmasıdır. İkinci sebebi ise bu dağılımların farklı parametre değerleri için sağa ve sola çarpıklık, basıklık, sivrilik vb. gibi özelliklerinin ortalamaları nasıl etkilediğini, hangi ortalamanın daha etkin olduğunu, birbirlerine olan üstünlük ve zayıflıklarını görmektir.

Normal dağılım biri konum (location) (μ) diğeri ölçek (scale) (σ^2) olmak üzere iki parametrelidir, reel değerlerden oluşan bir dağılımdır. Normal dağılımda ölçek parametresi dağılımın sivrilik ve basıklığını etkilemektedir. Bu nedenle Tablo 2 de normal dağılımın konum parametresi sabit tutulup $\mu = 0$, ölçek parametresi $\sigma^2 = 0.1; 1; 5$ değerleri ile değiştirilerek farklı veri sayılarında ($n = 10, 100, 1000$) çarpıklık ve basıklık değişiminin ortalamalar üzerine etkisi incelenmiştir. Yapılan simülasyon denemelerinden elde edilen ortalama ve varyans sonuçları Tablo 2 de verilmiştir. Bu sonuçlara göre, veri sayısı artışı ile $\sigma^2 = 0.1$ için gittikçe sivrileşen, $\sigma^2 = 1$ için standart normale yaklaşan ve $\sigma^2 = 5$ için de gittikçe basıklaşan bir dağılım oluşturacaktır. Oluşan dağılımlarda örneklem sayısı artıkça ortalamaların daha doğru ve güvenilir bir şekilde tahmin edilerek sonuçların sıfıra yaklaştıkları görülmektedir. Normal dağılımın simetrik ve belirli bir şekle sahip olması, ortalamaların belirli bir düzen içinde sıralanmasına neden olur. Bu durumda, $\bar{X}$ genellikle daha düşük bir değere sahip olurken, $\bar{B}$ ve M daha yüksek değerlerde elde edilmiştir. $n = 10, 100$ için $\sigma^2 = 0.1$ iken M değeri, $\bar{B}$ değerinden düşük elde edilirken diğer durumlarda $\bar{B}$ değeri M değerinden düşük kalmıştır. Düşük varyansa sahip bir ortalama, daha az belirsizlik ve

daha fazla güvenilirlik sunar. Bu anlamda varyanslar incelendiğinde $\bar{X}$ 'nın, $\bar{B}$ 'dan, $\bar{B}$ ise M 'dan daha etkin ve geçerli olduğu aşıkardır. Çünkü tüm varyans karşılaştırmaları $Var(\bar{X}) < Var(\bar{B}) < Var(M)$ olarak elde edilmiştir.

İki parametrelili gamma dağılımında şekil parametresi (k), bir konum parametresinin yaptığı gibi dağılım şeklini basitçe kaydırırken, bir ölçek parametresi (θ) gibi de dağılımı esnetip daraltabilmektedir. Bu sayede sağa çarpık bir dağılımın ortalamalar üzerindeki etkisini görmek mümkün olmaktadır. Dağılımın rassal değişkenleri $x \in (0, \infty)$ aralığında değer almaktadır. Tablo 3 de gamma dağılımının ölçek parametresi sabitken $\theta = 2$, şekil parametresinde k meydana gelen değişikliğin, değişen veri sayılarına da bağlı olarak ortalamalara etkisi incelenmiştir. Yapılan simülasyon denemelerinden elde edilen ortalama ve varyans sonuçları da Tablo 3 de verilmiştir. Sağa çarpık bir dağılımda M değerinin $\bar{X}$ değerinden küçük olacağı bilinmektedir. Örnekleme bağlı olarak çarpıklık ne denli büyüklük gösterirse gereken büyük örneklem hacminin gittikçe daha büyük olması gerekmektedir. Tablo 3 de θ parametresindeki artış ortalamaların artmasına sebebiyet verirken, $\bar{B}$ 'den küçük olan M değeri, veri sayısı arttıkça $\bar{B}$ değeri ile neredeyse aynı değere ulaşmıştır. $k = 1$ için en etkin ortalama $\bar{B}$ iken, $k = 2; 3$ için en etkin ortalama $\bar{X}$ dir. Bu durumun sebebi k değerinin artışının dağılımın sağa olan çarpıklığını sola doğru yansıtırmasından kaynaklanmaktadır. Aşırı sağa çarpık iken $\bar{B}$ en etkin iken, çarpıklığın azalması $\bar{X}$ 'nın daha etkin olmasına sebebiyet vermiştir. Bir diğer sonuç ise $\bar{B}$ 'nın tüm durumlarda M 'dan daha etkin olmasıdır. Bu durum $\bar{B}$ 'yı sağa çarpık dağılımlarda tüm verileri kullanmadan ölçüm veren M 'a göre üstün kılmaktadır.

Beta dağılımı, $[0,1]$ aralığında, iki parametre ile dağılımın şeklini kontrol eden, sonlu uzunluktaki aralıklarla sınırlı rastgele değişkenlerin davranışını modelleyen bir dağılımdır. Tablo 4 de şekil parametresi $\alpha = 5$ iken diğer şekil parametresi β değiştirilerek $[0,1]$ aralığında sola çarpık dağılımlar elde edilmiştir. Yapılan simülasyon denemelerinin ortalama ve varyans sonuçları incelenmiştir. Sola çarpık bir dağılımda genel olarak $\bar{X}$ 'nın M dan küçük olacağı bilinmektedir. Tablo 4'teki sonuçlara göre $\bar{B}$ değeri M 'dan küçük, $\bar{X}$ değerinden büyük elde edilmiştir. Etkinlik bakımından ise yine M kadar etkin ya da M dan daha etkin olduğu durumlar gözlemlenmektedir. $n = 1000$ ve $\beta = 1$ iken $Var(\bar{B}) = Var(M)$ olup, $n = 1000$ ve $\beta = 3$ iken $Var(\bar{X}) = Var(\bar{B})$ sonuçları elde edilmiştir. Bu sonuca göre $n = 1000$ iken tüm durumlarda varyanslar oldukça yakınlaşarak sonuçların eşitlendiği durumlar

gözlemlenmiştir. Bu durum yine $\bar{B}$ 'yı sola çarpık dağılımlarda tüm verileri kullanmadan ölçüm veren M 'a göre üstün kılmaktadır.

5.2. Veri Uygulamaları

Önerdiğimiz ortalamanın gerçek veri setlerindeki sonuçlarını ve tahmin varyansının değişimini anlamak amacıyla bu bölümde R Studio (version 4.4.1) programı kullanılarak uygulama yapılmıştır. Uygulamada hem farklı veri setleri hem farklı değişkenler hem de farklı gözlem sayıları kullanılarak binomik ortalamasının ($\bar{B}$) medyan (M) ve aritmetik ortalamaya ($\bar{X}$) göre ortalama ve tahmin varyansı bakımından nasıl şekillendiği incelenmiştir.

Örneklem medyanının varyans tahmini $Var(\hat{M})$ için Price ve Bonett (2001) tarafından önerilen McKean-Schrader tahmincisinin iki modifikasyonundan biri olan $V_{PB1} = \{(X_{(n-c+1)} - X_{(c)})/2z_{\alpha^*/2}\}^2$ kullanılmıştır. V_{PB2} kullanmamamızın sebebi V_{PB1} formülünden daha kompleks yapılı olmasıdır. Her iki modifikasyon da örneklem sayısı n büyüdükçe aynı değere yaklaşmaktadır. Örneklem aritmetik ortalamasının varyans tahminlemesi $Var(\hat{\bar{X}})$ için eşitlik (4.2.1) de verilen formül kullanılırken, örneklem binom ortalamasının varyans tahmini $Var(\hat{\bar{B}})$ için de eşitlik (4.2.2) kullanılmıştır.

Tablo 2. Farklı parametre değerlerine sahip normal dağılımların farklı gözlem değerlerinde 5000 örneklemin medyan, binom ortalama ve aritmetik ortalama sonuçlarının ortalama ve varyanslarının karşılaştırılması

	Parametreler	n	M	$\bar{B}$	$\bar{X}$	Ortalama Karşılaştırma	Var(M)	Var($\bar{B}$)	Var($\bar{X}$)	Varyans Karşılaştırma
Normal Dağılım	$\mu = 0$ $\sigma^2 = 0.1$	10	0,00243	0,00246	0,00166	$\bar{X} < M < \bar{B}$	0,01414	0,01194	0,01024	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		100	0,00053	0,00062	0,00047	$\bar{X} < M < \bar{B}$	0,00158	0,00145	0,00101	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		1000	0,00021	0,00019	0,00013	$\bar{X} < \bar{B} < M$	0,00016	0,00015	0,00010	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
	$\mu = 0$ $\sigma^2 = 1$	10	0,01081	0,00897	0,00581	$\bar{X} < \bar{B} < M$	0,14016	0,11739	0,09899	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		100	0,00257	0,00252	0,00196	$\bar{X} < \bar{B} < M$	0,01557	0,01431	0,01005	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		1000	0,00071	0,00068	0,00018	$\bar{X} < \bar{B} < M$	0,00161	0,00154	0,00100	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
	$\mu = 0$ $\sigma^2 = 5$	10	0,00637	0,00615	0,00588	$\bar{X} < \bar{B} < M$	0,69427	0,59295	0,51286	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		100	0,00327	0,00316	0,00075	$\bar{X} < \bar{B} < M$	0,08131	0,07365	0,05073	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		1000	0,00047	0,00044	0,00018	$\bar{X} < \bar{B} < M$	0,00783	0,00756	0,00500	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)

Tablo 3. Farklı parametre değerlerine sahip gamma dağılımlarının farklı gözlem değerlerinde 5000 örneklemin medyan, binom ortalama ve aritmetik ortalama sonuçlarının ortalama ve varyanslarının karşılaştırılması

	Parametreler	n	M	$\bar{B}$	$\bar{X}$	Ortalama Karşılaştırma	Var(M)	Var($\bar{B}$)	Var($\bar{X}$)	Varyans Karşılaştırma
Gamma Dağılım	$k = 1$ $\theta = 2$	10	1,48497	1,57915	1,99143	$M < \bar{B} < \bar{X}$	0,37785	0,33485	0,39601	Var($\bar{B}$) < Var(M) < Var($\bar{X}$)
		100	1,39581	1,40572	1,99842	$M < \bar{B} < \bar{X}$	0,04087	0,03715	0,04133	Var($\bar{B}$) < Var(M) < Var($\bar{X}$)
		1000	1,38762	1,38878	2,00157	$M < \bar{B} < \bar{X}$	0,00394	0,00383	0,00409	Var($\bar{B}$) < Var(M) < Var($\bar{X}$)
	$k = 2$ $\theta = 2$	10	3,45829	3,55962	4,00078	$M < \bar{B} < \bar{X}$	0,93199	0,80377	0,79424	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		100	3,36165	3,37276	3,99576	$M < \bar{B} < \bar{X}$	0,10080	0,09269	0,08087	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		1000	3,35696	3,35800	3,99898	$M < \bar{B} < \bar{X}$	0,01015	0,00987	0,00798	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
	$k = 3$ $\theta = 2$	10	5,45136	5,55397	5,99865	$M < \bar{B} < \bar{X}$	1,48718	1,25766	1,19261	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		100	5,35721	5,36818	6,00413	$M < \bar{B} < \bar{X}$	0,16679	0,15254	0,12161	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		1000	5,34915	5,35023	5,99825	$M < \bar{B} < \bar{X}$	0,01691	0,01647	0,01206	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)

Tablo 4. Farklı parametre değerlerine sahip beta dağılımlarının farklı gözlem değerlerinde 5000 örneklemin medyan, binom ortalama ve aritmetik ortalama sonuçlarının ortalama ve varyanslarının karşılaştırılması

	Parametreler	n	M	$\bar{B}$	$\bar{X}$	Ortalama Karşılaştırma	Var(M)	Var($\bar{B}$)	Var($\bar{X}$)	Varyans Karşılaştırma
Beta Dağılım	$\alpha = 5$ $\beta = 1$	10	0,86247	0,85651	0,83277	$\bar{X} < \bar{B} < M$	0,00282	0,00237	0,00204	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		100	0,86985	0,86913	0,83345	$\bar{X} < \bar{B} < M$	0,00031	0,00028	0,00020	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		1000	0,87051	0,87045	0,83332	$\bar{X} < \bar{B} < M$	0,00003	0,00003	0,00002	Var($\bar{X}$) < Var($\bar{B}$) = Var(M)
	$\alpha = 5$ $\beta = 2$	10	0,73231	0,72865	0,71462	$\bar{X} < \bar{B} < M$	0,00403	0,00333	0,00257	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		100	0,73419	0,73393	0,71383	$\bar{X} < \bar{B} < M$	0,00046	0,00042	0,00025	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		1000	0,73553	0,73550	0,71430	$\bar{X} < \bar{B} < M$	0,00005	0,00004	0,00003	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
	$\alpha = 5$ $\beta = 3$	10	0,63538	0,63337	0,62594	$\bar{X} < \bar{B} < M$	0,00408	0,00337	0,00260	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		100	0,63601	0,63582	0,62507	$\bar{X} < \bar{B} < M$	0,00048	0,00044	0,00027	Var($\bar{X}$) < Var($\bar{B}$) < Var(M)
		1000	0,63594	0,63591	0,62506	$\bar{X} < \bar{B} < M$	0,00005	0,00005	0,00003	Var($\bar{X}$) = Var($\bar{B}$) < Var(M)

Tablo 5. Farklı veri setlerine ait değişkenlerin medyan, binom ortalaması, aritmetik ortalama ve varyans tahminlerinin karşılaştırılması

Veri	Değişken	n	M	$\bar{B}$	$\bar{X}$	Ortalama Karşılaştırma	Var($\hat{M}$)	Var($\hat{B}$)	Var($\hat{X}$)	Varyans Tahminlerinin Karşılaştırması
BOD	Demand	6	15,800	15,331	14,833	$\bar{X} < \bar{B} < M$	7,12680	5,27691	3,57378	Var($\hat{X}$) < Var($\hat{B}$) < Var($\hat{M}$)
Dnase	Density	17	0,684	0,647	0,769	$\bar{B} < M < \bar{X}$	0,08372	0,05386	0,02264	Var($\hat{X}$) < Var($\hat{B}$) < Var($\hat{M}$)
	Conc	47	1,563	1,289	3,172	$\bar{B} < M < \bar{X}$	0,44310	1,40150	0,35944	Var($\hat{X}$) < Var($\hat{M}$) < Var($\hat{B}$)
Docvisits	Docvisit	100	2,000	1,786	3,330	$\bar{B} < M < \bar{X}$	0,06895	1,52511	0,26930	Var($\hat{M}$) < Var($\hat{X}$) < Var($\hat{B}$)
İris	Sepal.Length		5,800	5,789	5,843	$\bar{B} < M < \bar{X}$	0,00957	0,03167	0,00457	Var($\hat{X}$) < Var($\hat{M}$) < Var($\hat{B}$)
	Sepal.Width	150	3,000	3,009	3,057	$M < \bar{B} < \bar{X}$	0,00060	0,00877	0,00127	Var($\hat{M}$) < Var($\hat{X}$) < Var($\hat{B}$)
	Petal.Length		4,350	4,317	3,758	$\bar{X} < \bar{B} < M$	0,02152	0,14391	0,02078	Var($\hat{X}$) < Var($\hat{M}$) < Var($\hat{B}$)
Faithful	Eruptions	272	4,000	3,984	3,488	$\bar{X} < \bar{B} < M$	0,00503	0,04463	0,00479	Var($\hat{X}$) < Var($\hat{M}$) < Var($\hat{B}$)

$n = 6$ gözleme sahip BOD veri setinin Demand değişkeni (BOD_Demand) kullanılarak elde edilen $M, \bar{B}$ ve $\bar{X}$ ortalama sonuçları ve bu ortalamaların $Var(\hat{M}), Var(\hat{B})$ ve $Var(\hat{X})$ tahmin varyansları Tablo 5 de verilmiştir. BOD_Demand değişkeninin kullanılmasındaki amaç oldukça düşük gözlem sayısında ortalamaların ve varyans tahminlerinin değişimini incelemektir. Grafik 1(a) da verilen olasılık yoğunluk fonksiyonuna göre BOD_Demand değişkeni normal olmayan, hafif sola çarpık bir dağılıma sahiptir. Sola çarpık dağılımlarda $\bar{X}$, dağılımın sol tarafındaki düşük değerlerden etkilenir ve bu nedenle M 'dan daha düşük bir değere sahip olma eğilimindedir. Elde edilen sonuçlara göre de $\bar{X}$, M 'dan hem ortalama bakımından hem de tahmin varyansı bakımından düşük elde edilmiştir. $\bar{B}$ ise hem ortalama hem de tahmin varyansı açısından M 'dan küçük, $\bar{X}$ 'dan ise büyük elde edilmiştir. Bu oldukça düşük gözlem değerlerinde $\bar{B}$ 'nın, M 'dan daha etkin bir ortalama olduğunu göstermektedir.

DNase veri setinin gözlem sayısı $n = 176$ olup Density değişkeni (DNase_Density) (0,2) aralığında değişen değerler almaktadır. Veri setinin [160,176] aralığındaki $n = 17$ adet gözlemi alınarak DNase_Density değişkeni için örneklem oluşturulmuştur. Bu şekilde spesifik bir örneklem belirleme ihtiyacımız Grafik 1(a) da verilen dağılımın tam tersi olduğu durumda $M, \bar{B}$ ve $\bar{X}$ ortalamalarının ve bu ortalamaların $Var(\hat{M}), Var(\hat{B})$ ve $Var(\hat{X})$ tahmin varyanslarının nasıl değiştiğini kıyaslayabilmektir. Bu doğrultuda Grafik 1(b) de verilen olasılık yoğunluk fonksiyonuna göre DNase_Density değişkeni normal olmayan, hafif sağa çarpık bir dağılıma sahiptir. Böyle bir durumda genel olarak M , dağılımın merkezini temsil ederken, $\bar{X}$ sağdaki daha büyük değerlerden etkilenerek daha yüksek bir değere ulaşır. Tablo 5 de elde edilen sonuçlar bu durumu desteklerken, $\bar{B}$ 'nın M 'dan daha küçük bir ortalama ve tahmin varyansı değerine sahip olduğu görülmektedir. Bu doğrultuda $\bar{B}$ 'nın $\bar{X}$ ve M 'na göre daha iyi bir tahmin ortalaması ve varyansına sahip olup bu tür bir örneklemi daha iyi temsil edeceği sonucuna ulaşılır.

Conc değişkeni (DNase_Conc), DNase veri setinin gözlem değerlerinde sıfıra yakın değerlerin yoğunlukta olduğu ve aykırı değerler de içeren bir diğer değişkenidir. Yine

bu değişkenin [130,176] aralığındaki $n = 47$ adet gözlem değeri için örneklem oluşturulmuştur. Grafik 1(c) de elde edilen olasılık dağılımının Grafik 1(b) de elde edilen dağılıma göre daha sağa çarpık olup, aşırı sağa çarpık bir dağılım sergilediği görülmektedir. Böyle bir dağılım için elde edilen $M, \bar{B}$ ve $\bar{X}$ ortalama sonuçları ve bu ortalamaların $Var(\hat{M}), Var(\hat{B})$ ve $Var(\hat{X})$ tahmin varyansları Tablo 5 de verilmiştir. Aşırı sağ çarpık dağılımlarda, sağ tarafta çok sayıda uç değer (outlier) bulunur. Bu uç değerler, dağılımın genel yapısını bozarak, $\bar{X}$ ve M arasındaki farkı artırır. Böyle bir çarpıklıkta $\bar{B}$ daha düşük bir ortalama ile uç değerlerin etkisini oldukça elemine ettiği açıktır. Ancak tahmin varyanslarına bakıldığında aynı durum söz konusu değildir. Aşırı sağa çarpık bir dağılımda, ortalamanın varyans tahmininin medyanın tahmin varyansından küçük olması beklenmez. Genellikle, ortalama daha yüksek bir değere sahip olduğu için, varyansı da daha büyük olacaktır. Ancak, belirli durumlarda ve özel veri setlerinde bu tür bir durum gözlemlenebilir. Bu veri seti bu anlamda iyi bir örnektir. Uç değerlerin var olduğu durumlarda varyansın genellikle yükselmesi beklenir ancak etkin bir ortalama da düşük kalması istenir. Bu değişken için $Var(\hat{B}) = 1,40150$ sonucu dağılımdaki yayılımdan varyansın etkilendiğini göstermektedir.

R programının “zic” paketinde bulunan Docvisits veri setinde $n = 1812$ adet gözlem bulunmaktadır. Veri setindeki Docvisits değişkeninde (Docvisits_Docvisits) 0 ve 1 değeri oldukça fazladır ve aykırı değerlerde bulunmaktadır. Bu değişkenin ilk $n = 100$ gözlemi alınarak $M, \bar{B}$ ve $\bar{X}$ ortalama sonuçları hesaplanmış ve bu ortalamaların $Var(\hat{M}), Var(\hat{B})$ ve $Var(\hat{X})$ tahmin varyansları Tablo 5 de verilmiştir. Değişkenin olasılık dağılımı Grafik 1(d) de verilmiştir. Grafik 1(c) gibi aşırı sağa çarpık bir dağılıma sahip olup, yine Şekil 1(c) den daha sola dayalı ve sivri bir görüntü vermektedir. $\bar{B}$, DNase_Conc değişkeninde olduğu gibi bu değişkende de yine en küçük iken ortalamaların varyans tahminlerinde durum değişmiştir. En küçük tahmin varyansı M 'da, ardından $\bar{X}$ 'dadır. Bu değişkende de $Var(\hat{B})$ en yüksek çıkmıştır. $\bar{B}$, sağdaki uç değerlerden ortalama bakımından etkilenmez iken tahmin varyansı bakımından fazla etkilenmektedir. Bu durumun bir başka sebebi ise $\bar{B}$ 'nın varyans tahminindeki hesaplamının veri sayısının artışından etkilenmesidir. $Var(\hat{B})$ formülündeki kombinasyon hesabı yüksek gözlemlerde oldukça artarken, $\bar{X}$ ve M 'nın varyans

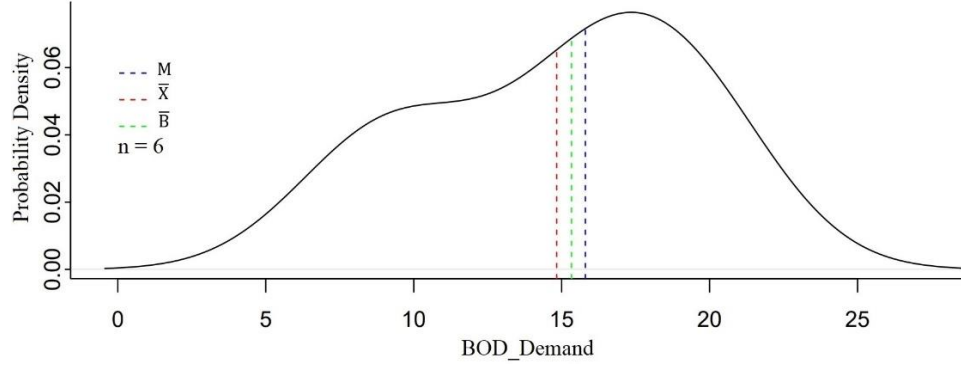
tahmininde tam tersi söz konusudur. Veri sayısı arttıkça tahmin varyansı küçülmektedir. Elbette veri sayısı arttıkça sapmaların azalmasını bekleriz ama bu durum her veri seti için mümkün değildir. Bu durumun bir başka sonucu ise $\bar{B}$ 'nın aşırı sağa çarpık dağılımlara düşük gözlem sayılarında hem ortalama hem de tahmin varyansı açısından iyi sonuçlar vermesidir.

$n = 150$ gözlem değerine sahip İris veri setinin Sepal.Length (İris_Sepal.Length), Sepal. Width (İris_Sepal.Width) ve Petal.Length (İris_Petal.Length) değişkenleri kullanılarak elde edilen $M, \bar{B}$ ve $\bar{X}$ ortalama sonuçları ve bu ortalamaların $Var(\hat{M}), Var(\hat{B})$ ve $Var(\hat{X})$ tahmin varyansları Tablo 5 de verilmiştir. Bu veri setindeki üç değişkeni seçmemizin sebebi İris_Sepal.Length değişkeninin basık bir normal dağılıma, İris_Sepal.Width değişkeninin sivri bir normal dağılıma ve İris_Petal.Length değişkeninin ise çift modlu bir dağılıma (verilerin iki farklı grup veya süreçten geldiğini gösterir) sahip olmasıdır. Öyle ki, İris_Sepal.Length, İris_Sepal.Width ve İris_Petal.Length değişkenlerinin olasılık fonksiyonlarının dağılımları sırasıyla Grafik 1(e), Grafik 1(f) ve Grafik 1(g) de verilmiştir. Basık ya da sivri bir normal dağılımda ortalamaların birbirlerine yakın çıkmaları beklenir. Çünkü simetrik bir normal dağılımda aritmetik ortalama medyana eşittir. İris_Sepal.Length ve İris_Sepal.Width değişkenlerinin ortalamalar bakımından oldukça yakın değerler aldıkları, İris_Petal.Length değişkeninde ise $\bar{X}$ 'nın M ve $\bar{B}$ 'ya göre daha düşük çıktığı görülmüştür. Grafik 1(g) de dağılımın dalgalanması incelendiğinde ortalama olarak binom ve medyanın daha merkezci bir sonuç verdiği görülmektedir. Varyans tahminleri bakımında üç değişken için de yapılabilecek ortak çıkarım binom tahmin varyansının her bir değişkende diğer tahmin varyanslarına göre yüksek çıkmasıdır. Bu tür düzgün yapıli dağılımlara sahip değişkenlerde binom tahmin varyansının bu denli yüksek çıkmasının sebebi yine gözlem sayısı ile alakalı bir durumdur. Çarpıklığın etkisinde olmayan bir dağılımda bile binom tahmin varyansının bu denli diğerlerinden ayrılması yine gözlem sayısı ile ilişkilendirilebilir.

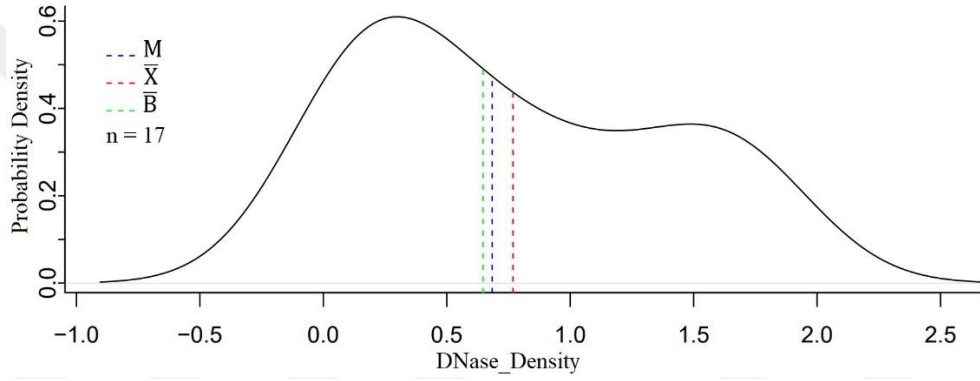
Faithful veri setinin gözlem sayısı $n = 272$ olup Eruptions değişkeni (Faithful_Eruptions) kullanılarak elde edilen $M, \bar{B}$ ve $\bar{X}$ ortalama sonuçları ve bu ortalamaların $Var(\hat{M}), Var(\hat{B})$ ve $Var(\hat{X})$ tahmin varyansları Tablo 5 de, olasılık

dağılımı ise Grafik 1(h) de verilmiştir. Bu değişken İris_Petal.Length değişkeni ile bir kıyaslama yapabilmek adına verilmiştir. Değişkenin dağılımı tıpkı İris_Petal.Length değişkeninde olduğu gibi çift modlu bir dağılım sergilemektedir ve gözlem sayısı daha fazladır. Bu durum farklı veri setlerindeki benzer dağılıma sahip değişkenleri incelememize ve sonuçlarımızın ne denli tutarlı olduğunu anlamaya olanak tanımaktadır. Elde edilen sonuçlar kıyaslandığında burada da $\bar{X}$ 'nın M ve $\bar{B}$ 'ya göre daha düşük çıktığı görülmüştür. Varyans tahminleri açısından da yine İris_Petal.Length değişkenindeki duruma benzer bir sonuç çıkmıştır. Bu sonuç gözlem sayısındaki artışın çift modlu dağılıma sahip değişkenlerde ortalama ve tahmin varyansı sıralamalarını değiştirmedigini göstermektedir.

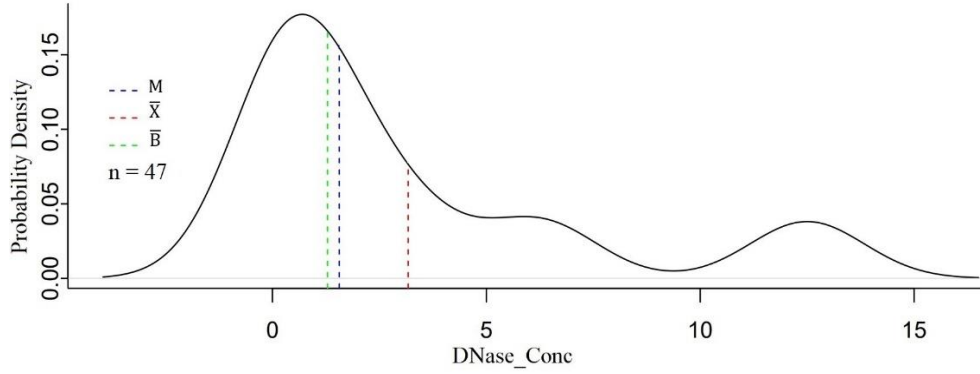
Grafik 1. Farklı veri setlerine ait değişkenlerin olasılık yoğunluk fonksiyonları ve medyan, binom ortalaması ve aritmetik ortalama sonuçlarının karşılaştırılması



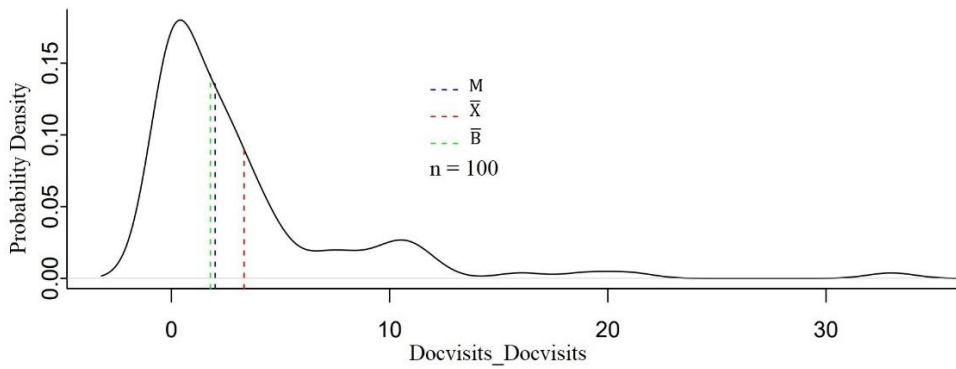
(a)



(b)

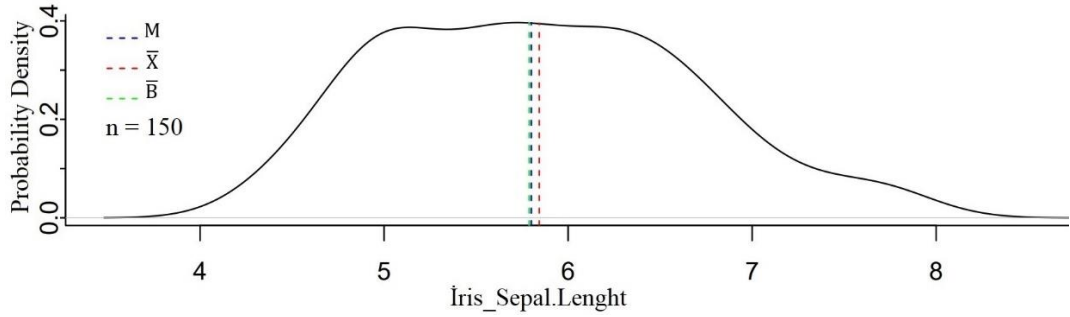


(c)

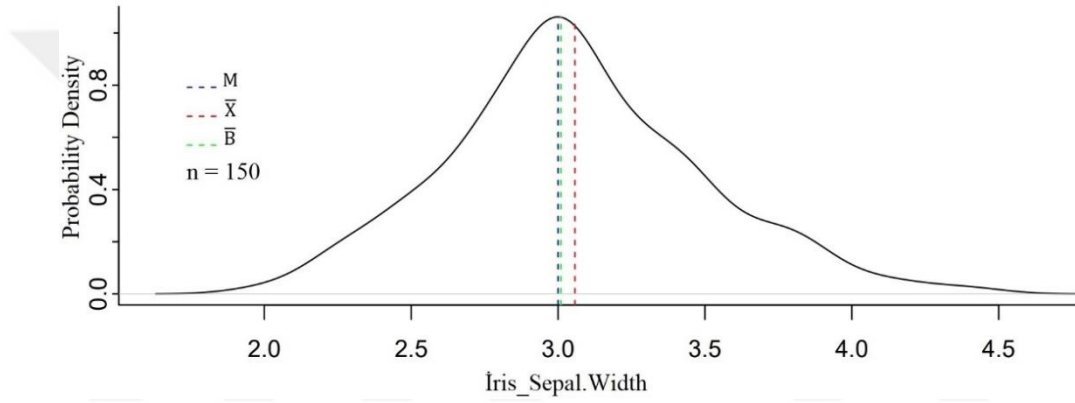


(d)

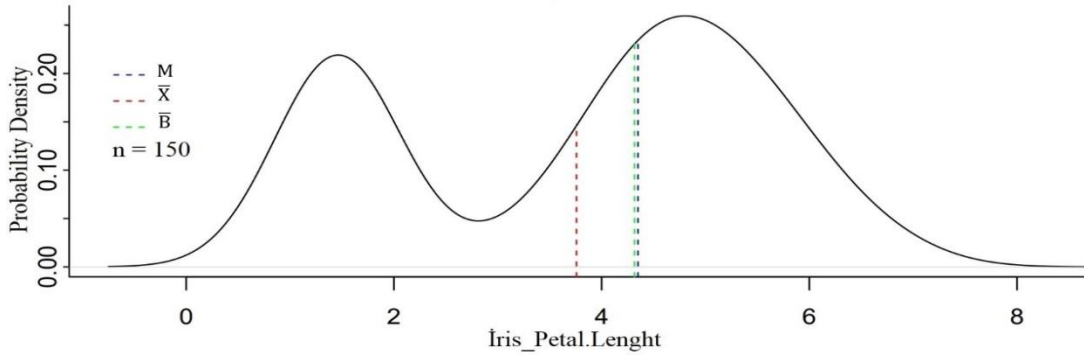
Grafik 1. Farklı veri setlerine ait değişkenlerin olasılık yoğunluk fonksiyonları ve medyan, binom ortalaması ve aritmetik ortalama sonuçlarının karşılaştırılması



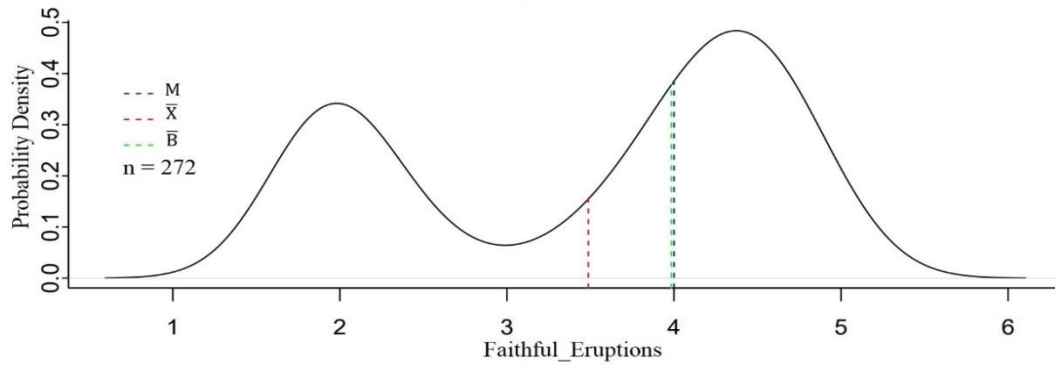
(e)



(f)



(g)



(h)

SONUÇ

İstatistiksel anlamda ortalama, bir verinin merkez noktasını temsil etmektedir. Bu temsili, parametrik ortalamalardan aritmetik, kareli, geometrik, harmonik vb. parametrik olmayan ortalamalardan ise mod, medyan vb. ortalamalar yapabilmektedir. Literatürdeki ortalamaların birbirlerine üstün ya da zayıf olduğu durumları elemine etmek, dezavantaj yaratan eksiklikleri gidermek için, bir geliştirme çalışmasının sonucu olarak binomik ortalama önerilmiştir.

Binomik ortalama tıpkı aritmetik ortalamada olduğu gibi tüm gözlem değerlerini hesaba katarak bir hesaplama yapar. Bu yönü ile binomik ortalama parametrik bir ortalama davranışı sergiler. Bilindiği üzere aritmetik ortalama aykırı değerlere karşı son derece duyarlıdır. Aykırı değer var olduğu durumlarda mod, medyan ve kartiller vb. parametrik olmayan ortalamalardan ya da winsorize ortalama, budanmış (trimmed) ortalama, M-tahmincileri vb. dayanıklı (robust) ortalama ölçülerinden kullanılmalıdır. Bu ölçümlerin binomik ortalama ile benzerlik gösteren yanı ise gözlem değerlerinin küçükten büyüğe sıralanarak hesaplama yapılmasıdır. Bu yönü ile de binomik ortalama parametrik olmayan bir ortalama davranışı sergiler. Bu özelliği ile binomik ortalama, istatistik alanında yapılacak çalışmalarda veri setlerinin yorumlanması yapılırken parametrik ve parametrik olmayan ortalamalara alternatif olarak yarı parametrik bir ortalama gibi kullanılabilir.

Binomik ortalama, ağırlıklı ortalama mantığına benzerdir, ancak uygulaması açısından oldukça farklılık arz etmektedir. Çünkü ağırlıklı ortalamada ağırlıklar gözlem sayısına bağlı olarak belirlenmez. Hangi gözlem sayısında ağırlığın ne olacağı belli değildir. Fakat binomik ortalama da her bir gözleme gelecek ağırlık binom katsayıları ile önceden bellidir. Sıfır ve negatif değerlere sahip veri kümeleri için, geometrik ve harmonik ortalama kullanılamamasına rağmen binomik ortalamada tıpkı aritmetik ortalamada olduğu gibi tüm reel yani sayısal değerler kullanılabilir. Binom katsayıları uç değerlerden ortaya doğru büyüdüğünden, en büyük ağırlığı ortaya denk gelen gözlem değeri almaktadır. Bu durum verinin merkezindeki gözlemin etkisini oldukça arttırırken, uç kısımlarda kalan verilerin etkisini de oldukça azaltmaktadır. Bu

anlamda da budanmış ortalama da veri kaybı olurken, binomik ortalama da veri kaybı olmadan aşırı uçlardaki verinin etki değeri azaltılmaktadır.

Çalışmanın normal dağılım ile ilgili simülasyonlarında genel anlamda hem ortalama olarak hem de varyanslar açısından M 'dan daha etkin çıkmıştır. Gamma dağılımında ise genel anlamda ortalama olarak $\bar{X}$ 'dan daha düşük sonuçlar verirken, varyans olarak M 'dan daha düşük sonuçlar vermiştir. Beta dağılımında ise yine hem ortalama olarak hem de varyanslar açısından M 'dan düşük değerler vermiştir. Bu simülasyon çalışması tüm ortalamaları tek bir kriter altında incelemek açısından oldukça önemlidir. Çünkü 5000 adet örneklemden her bir ortalama için elde edilen sonuçların hepsi ortalama ve varyans çatısı altında birleştirilmiştir.

Gerçek veri uygulaması ile yapılan çalışmadan elde edilen sonuçlara göre binom ortalamasının avantaj ve dezavantajları şu şekilde sıralanabilir:

- İstatistiksel olarak yüksek hacimli örneklerde ortalamalar genellikle birbirlerine yakınsama eğilimi gösterirler. Hatta aşırı çarpıklık içeren örneklerde gözlem ekleme yoluna gideriz. Bu nedenle asıl mesele düşük hacimli örneklerde iyi sonuçlar elde etmektir. Bu nedenle $\bar{B}$, düşük gözlem sayılarında diğer ortalamalara göre ortalama olarak düşük sonuçlar verirken, tahmin varyansı bakımından M 'dan daha düşük sonuçlar vermiştir.
- Veri küçükten büyüğe sıralandığında uçlardaki aykırı değerleri elemine etmekte oldukça başarılıdır. Çünkü binom katsayılarının değeri uçlardan merkeze doğru artmaktadır. Bunun sonucu olarak da hiçbir veri kaybına olanak vermeden aykırı değerler ile M 'dan daha etkin sonuçlar elde edilebilmektedir.
- Aşırı çarpık dağılımlarda ortalama olarak düşük sonuçlar verirken varyans tahmini açısından düşük sonuçlar vermemektedir. Bu durum varyans tahmin formülünün pay kısmında kalan kombinasyon hesabından kaynaklanmaktadır. Gözlem sayısındaki artışa payın verdiği cevabı payda karşılayamamaktadır. Bu nedenle bu ortalamanın tahmin varyansına yönelik çalışmalar geliştirilebilir.

- Çok yüksek gözlem sayısına sahip veri setlerinde ($n > 1000$) ortalama ve tahmin varyansındaki kombinasyon hesaplarını yapmaya programların mevcut işlemci gücü yetmemektedir.

Hem aritmetik ortalamanın hem de binomik ortalamanın tüm gözlem değerlerini kullanarak hesaplama yaptığı düşünüldüğünde, binomik ortalamanın matematiğin salınım teorisi, regresyon analizi, robust teknikler vb. alanlarda kullanımına yönelik çalışmalar yapılabilir. Yine medyan değeri ile kıyaslandığında ise neredeyse tüm sonuçlarda medyana yakın değerler verip, hem de tüm gözlemleri kullanmasından dolayı da medyan yerine bir alternatif oluşturabilir.



KAYNAKÇA

- Akdi, Y. (2011). *Matematiksel İstatistiğe Giriş*. Ankara: Gazi Kitabevi.
- Bakker, A., Gravemeijer, K. (2006). An historical phenomenology of mean and median. *Educational Studies in Mathematics*, 149-168.
- Beirne, P. (1987). Adolphe Quetelet and the origins of positivist criminology. *American Journal of Sociology*, 1140-1169.
- Byun, S. H., Poznanovic, S. (2024). On the maximum of the weighted binomial sum $(1 + \alpha)^{-r} \sum_{i=0}^m \binom{m}{i} a^i$. *Discrete Mathematics*, 113925.
- Caillat, J. (1923). La Methode scientifique selon Pascal. Formation et apprentissage de la Methode. *Revue d'Histoire litteraire de la France*, 129-162.
- Conover, W. J. (1999). *Practical nonparametric statistics*. John Wiley and sons.
- Fisher, R. A. (1936). Statistical methods for research workers. *Statistical methods for research workers*.
- Glasby, S. P., Paseman, G. (2021). On the maximum of the weighted binomial sum $2^{-r} \sum_{i=0}^m \binom{m}{i} a^i$. *arXiv preprint arXiv:2109.11396*.
- Hald, A. (1998). A History of Mathematical Statistics from 1750 to 1930.
- Heath, T. L. (1921). *A history of Greek mathematics*. Oxford University Press.
- Heath, T. L. (1956). *The thirteen books of Euclid's Elements*. Courier Corporation.
- Hogg, R. V., McKean, J. W., & Craig, A. T. (2013). *Introduction to Mathematical Statistics*. India: Pearson Education.
- Karacasu, M., Yayla, N. (2004). Kentiçi Otobüs Taşımacılığında Özelleştirme İçin Bir Karar Destek Modeli Önerisi:Eskişehir Örneği. *İtÜ Dergisi*.
- Mosteller, F. (1964). Samuel S. Wilks: statesman of statistics. *The American Statistician*, 11-17.
- Pearson, E. S. (1936). Karl Pearson: An appreciation of some aspects of his life and work. *Biometrika*, 193-257.
- Pearson, K. (1895). X. Contributions to the mathematical theory of evolution. 2. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London*, 343-414.
- Pfanzagl, J. (2011). *Parametric statistical theory*. Walter de Gruyter.

- Price, R. M., Bonett, D. G. (2001). Estimating the Variance of Sample Median. *Journal of Statistical Computation and Simulation*, 295-305.
- Sanger, C. (1901). *AL Bowley. Elements of Statistics*. England: Oxford University Press Oxford.
- Scott, J. F. (1936). John Wallis as a historian of mathematics. *Annals of Science*, 335-357.
- Sprent, P., Smeeton, N. C. (2016). *Applied nonparametric statistical methods*. CRC press.
- Struik, D. (1958). Historically Speaking, Omar Khayyam, mathematician. *The mathematics teacher*, 280-285.
- Waerden, B. L. (2013). *A history of algebra: From al-Khwarizmi to Emmy Noether*. Springer Science and Business Media.
- Wilson, R. (2016). *Combinatorics: A very short introduction*. Oxford University Press.