

## İnternet Sitesi İstatistiklerinin Orta Vadede Değişimi

Gözde BOZKURT\*, Ahmet Mete ÇİLİNGİRTÜRK\*\*

### Öz

Bilgisayar ağlarının planlanması ve kaliteli ağ servisi sağlamak için internet veri trafiği modellerinin tahminlenmesine ihtiyaç duyulmaktadır. Bu çalışmada amaç, tıklama verilerinden meydana gelen sayfa merkezli verilere ait dağılımların modellenmesi ile benzetim ve planlama modellerine kaynak oluşturmaktır. Bu amaç doğrultusunda hedef olarak belirlenen web siteleri, aksesuar ve hazır giyimin pazarlandığı e-ticaret ile bilgi edinme amacı taşıyan haber siteleridir. Bu web sitelerinin göstergelerinden olan hemen çıkma oranı, sitede kalma süresi ve sayfa görüntüleme sayısına ait dağılımlar incelenmiştir. Beş ay aralıkla elde edilen verilerin karşılaştırmalı analiz edilmesiyle, hemen çıkma oranı ve sayfa görüntüleme sayısı göstergelerinin, yapılacak benzetim modellerinde Dagum ve Johnson SB dağılımları ile incelenebilir olduğu çıkarımı yapılabilmektedir. İnternet kullanıcılarının davranış şeklinin değişmesi nedeniyle, araştırmacıların diğer web sitesi göstergeleri için sürekli güncel verilere dayalı dağılımları kullanarak değerlendirmeler yapmaları uygun olacaktır. Elde edilen dağılım bilgileriyle, özellikle özel sektörde kullanılan satış simülasyonlarında, web siteleri arasındaki performans değerlendirmesi gibi çeşitli alanlarda kullanılabilceği düşünülmektedir.

**Anahtar Kelimeler:** Benzetim Modelleri, Web Sitesi İstatistikleri, İstatistiksel Dağılımlar, İnternet Kullanım Davranışı, Orta Vade Değişim

**JEL Kodu:** C46, L86, M31

### Change in the Medium Term of Site-centric Web Statistics

### Abstract

Internet data traffic models need to be estimated to provide computer network planning and high quality network service. The aim of this study is to model the distributions of page-based data consisting of click data, and to generate resources for simulation and planning models. For this purpose, web sites, accessories and ready-to-



**Özgün Araştırma Makalesi (Original Research Article)**

**Geliş/Received:** 13.09.2019

**Kabul/Accepted:** 25.10.2019

**DOI:** <http://dx.doi.org/10.17336/igushbd.619656>

\* Sorumlu Yazar: Doktora Öğrencisi, Marmara Üniversitesi, Ekonometri Bölümü, İstanbul, Türkiye, E-posta: [gozdebozkurt92@gmail.com](mailto:gozdebozkurt92@gmail.com), **ORCID ID** <https://orcid.org/0000-0001-8413-1099>

\*\* Prof. Dr., Marmara Üniversitesi, Ekonometri Bölümü, İstanbul, Türkiye,

E-posta: [acilingi@marmara.edu.tr](mailto:acilingi@marmara.edu.tr), **ORCID ID** <https://orcid.org/0000-0001-8677-7969>

wear websites are presented as e-commerce and news sites aiming to acquire information. Bypass rate, duration of site and number of page views are examined. With comparative analysis of the data obtained at five-month intervals, it can be concluded that the bounce rate and page video number indicators are examined by Dagum and Johnson SB distributions in the simulation models to be made. Due to the change in the behavior of Internet users, it would be appropriate for the researchers to make evaluations using other distributions based on current data for other website indicators. It is thought that the distribution information obtained can be used in various areas such as performance evaluation among web sites, especially in sales simulations used in the private sector.

**Keywords:** Simulation Models, Web Site Statistics, Statistical Distributions, Internet Usage Behavior, Medium Term Change

**JEL Code:** C46, L86, M31

## 1. Giriş

Yüksek ivmeli gelişime sahip olan teknoloji, iletişim başta olmak üzere pek çok farklı disiplinde etkisini yoğun bir şekilde hissettirmektedir. Gelişen teknolojinin ürünü olan internet pazarlama fonksiyonunda kurumlar ile müşteriler arasındaki geleneksel iletişim kanallarını güncellemiştir. İletişimi yeni kanallara taşınarak daha büyük bir iletişim ağı geliştirmiştir. Sosyal yazılımlar olarak ifade edilen, internetle birlikte meydana gelen etkileşimli uygulamalarla birlikte bu gelişim ivmelenmiştir. Kaliteli ağ servis hizmetlerinin sağlanması ve planlanması adına internet tabanlı veri trafiğinin de tahmin edilmesine ihtiyaç duyulmaktadır. Benzetim ve planlama modelleri, gerçek hayattaki olgulardan yola çıkılarak deneysel sonuçların üretilmesine olanak tanımaktadır. Bu modellerin ele alındığı çalışmalarda genellikle internet tabanlı web sitelerine ait kullanıcı merkezli tıklama verileri kullanılmaktadır.

Bu çalışmanın amacı, bahsedilen tıklama verilerinden meydana gelen sayfa merkezli verilere ait dağılımların tahmin edilerek benzetim ve planlama modellerine kaynak oluşturmaktır. Bu çalışmada aksesuar ve hazır giyim pazarlandığı e-ticaret siteleri ve bilgi edinimi amacı taşıyan haber siteleri hedeflenmiştir. Farklı yapıdaki bu web sitelerinin içerikleri nedeniyle web sitesinde bulunma süresi ve kullanıcıların gezdikleri sayfa sayıları bakımından ayrıştığı veya bunun aksine farklılık göstermediği yönünde iki farklı görüş mevcuttur. Çalışmanın ana hedefi, kısa dönemde web sitesi istatistiklerinin dağılım şekillerinin değişip değişmediğini belirleyerek, planlamada belli dağılımlar üzerinde durulmasını önermektir.

## 2. Literatür Taraması

Planlama ve benzetim modellerine ait ilgili literatüre bakıldığında modellere ait iki farklı teori yer almaktadır. Bunlardan ilki sosyal etki teorisidir. Sosyal bir durum karşısında bireylerin birbirilerinin etkilemesi üzerinde durulmaktadır. Sosyal etki, kaynakların algılanan etki gücü, kaynağa tepki veren birey sayısı ve meydana gelen olayın anlık duyurulmasının fonksiyonu olarak meydana gelmektedir (Levina ve Vilnai-Yavetz, 2015). Diğer bir yaklaşım ise sosyal mevcudiyet teorisidir. İletişim ortakları arasındaki fiziksel, görsel veya işitsel iletişimin sağlanabildiği ölçüde medya kabul edilmektedir. Bu doğrultuda da, iletişim ortakları arasındaki davranış etkileşimi sosyal mevcudiyetle doğru orantılı yükseliş gösterecektir.

Web sunucuları sayesinde geliştirilen ve iletişim ortakları arasındaki iletişim sağlandığı kanallardan biri olan e-ticaret sitelerinin performans ölçümlerinde tıklama, tıklama başına ciro ve hit kullanılmaktadır. Mevcut araştırmalarda yer alan hit miktarının, 1995'te %2 ve 2010'a gelindiğinde ise %0,08 oranları ile satın alma davranışına dönüştüğü ortaya konulmuştur. Bu nedenle de yöneticiler tarafından, tıklama adedi ile satın alma oranının yetersiz ve yanlış yönlendirici bilgi olduğunun düşünülmüştür. Dolayısı ile web sitelerine girişin ve geçirilen sürenin daha uygun göstergeler olduğu savunulmuştur (Dalessandro vd., 2012). Bunun yanı sıra bir web sitesinde geçirilen sürenin azalmasıyla siteye daha sık giriş yapıldığı belirlenmiştir (Johnson vd., 2003). Göstergelere göre incelenmesi hedeflenen web sitelerine ait "kullanıcı" ve "site" merkezli verileri elde etmek mümkündür. Ancak kullanıcı merkezli verilerin, bireysel kayıt ve ortak bilgisayar kullanımı gibi bilgilere ihtiyaç duyması, bu veri toplama tekniğinin dezavantajını oluşturmaktadır (Bucklin ve Sismeiro, 2009). Web sitelerinin göstergelere ait verilerinin elde edilmesi sonucu kullanıcıların site kullanım davranışları incelenildiğinden, reklam ve ikna amacı olarak kullanılan internetin rolü ve etkisi çevrimiçi pazarlama davranışlarının da anlaşılabilirliğine olanak tanımaktadır. Kişi bazlı verilerin kullanılması ile internet satın alma davranışının zamana göre trendi belirlenebilmektedir. Ayrıca tıklama dağılımının bilinmesi durumunda öngörü yapılmasının mümkün olduğu belirtilmiştir. İlgili sayfaya yapılan önceki giriş sayılarının ve sayfaya girdikten sonra geçirilen sürenin modellerde anlamlı olduğu görülmüştür. Aynı zamanda modellere ait parametrelerin Gaussian ve birim toplamlarının ise Multinomial dağılım gösterdiği kabul edilmiştir (Nottorf, 2014). Kullanıcı bazlı internet verisinin hacmi veya indirilmesi ile yükleme miktarlarına ait dağılımların günlük bazlı incelenmesinde, bu dağılımların Weibull, Lojistik, Rician ve Nakagami dağılımlarına genellikle uyum göstereceği öngörüsü doğrultusunda model tahminleri gerçekleştirilmiştir. Elde edilen sonuçta ise parametrik olmayan dağılımlara en iyi uygunluk gösterdiği ve bunun yanı sıra da en düşük log-benzerlik değerinin Nakagami dağılımında olduğu tespit edilmiştir (Adeyemi vd., 2018). Bir web sayfasından diğerine geçiş yapan kullanıcıların, dolaştığı sayfa sayısı göstergesinin Ters Gaussian dağılıma uyduğu belirlenmiştir. Böylece web sayfalarının kurulma maliyetleri ile tasarımlarının modellenmesinde faydalı bir bilgi olarak kullanılabilceği ileri sürülmüştür. Bunun yanı sıra, bazı firmalar kullanıcıların daha fazla sayfayı ziyaret etmesini sağlamak amacıyla bilgileri daha fazla sayfaya yaymaktadır. Bu nedenle, sayfa sayısının yerine sayfada geçirilen toplam sürenin dikkate alınması gerekliliği de vurgulanmıştır (Huberman vd., 1998). Hazır giyim alanında e-ticaret gerçekleştiren firmalarının ele alındığı bir çalışmada, bu alana ait dört ürüne ait işlemler kullanılarak firmaların ürün bağımlılığına dayalı rekabet durumları açıklanmaya çalışılmıştır. Elde edilen sonuçta, bu sektörü ulusal ve özel olmak üzere toplamda 10 markanın temsil ettiği görülmüştür (Heuer vd., 2015).

### 3. Çalışma Metodu

Çalışma amacına uygun olarak incelenmesi hedeflenen veriler, web sitelerine ait ölçümlerin yapıldığı ve yayınlandığı "Alexa Traffic Rank" sitesi aracılığıyla temin edilmiştir. 62 adet haber ve 65 adet moda web sitesine ait veriler derlenmiştir. Ana sayfadan ayrılma oranı (bounce rate), aynı sitede ortalama açtıkları sayfa sayısını ifade eden sayfa görüntüleme sayısı (daily pageviews per visitor) ve ziyaretçilerin bu sitede ortalama gezinme sürelerini ifade eden sitede kalma süresi (daily time) göstergeler kurama dayalı olarak seçilmiştir. Beş ay aralıklı Ağustos 2018 ve Ocak 2019 site istatistiklerinin dağılımları tahmin edilerek, karşılaştırma amaçlanmıştır. Güncel tarihli olan web sitesi istatistikleri, moda ve haber siteleri olacak şekilde gruplandırılmıştır.

Gruplar arası farklılık durumunun incelenmesi amacıyla t-testi uygulanmıştır. Gruplar arası anlamlı fark görüldüğünden, bu iki farklı kaynağa ait verilerin dağılımları ayrı ele alınmıştır. Analizler EasyFit 5.6 ve SPSS-18 versiyonları kullanılarak yürütülmüştür.

#### 4. Bulgular

Bu çalışmanın ana amacı, web sitesi odaklı internet ölçülerine ait uygun dağılımların incelenmesidir. Bu doğrultuda öncelikle, haber ve e-ticaret sitelerine göre incelenen değişkenlerde anlamlı farklılık olup olmadığı sınanmıştır. Sonuçlar **Tablo.1'**de özetlenmiştir.

**Tablo.1:** Değişkenlere Ait İstatistiksel Analiz Sonuçları

Değişken Analiz	Sitede Kalma Süresi		Hemen Çıkma Oranı		Sayfa Görüntüleme Sayısı	
	Ağustos 2018	Ocak 2019	Ağustos 2018	Ocak 2019	Ağustos 2018	Ocak 2019
Giyim-E-Ticaret	05:37 (03:42)	06:22 (04:37)	32,8046 (10,22508)	43,1231 (20,34036)	5,3035 (2,66857)	4,5903 (2,83975)
Eğiklik/Basıklık	2,170/5,779	1,218/2,052	1,005/1,150	1,123/0,474	2,377/10,309	1,862/6,527
Haber-Kaynak	06:14 (04:01)	08:23 (06:13)	54,7954 (14,25277)	64,0129 (17,54191)	3,1128 (1,85145)	2,4798 (1,45565)
Eğiklik/Basıklık	1,617/2,903	1,120/1,599	-0,022/- 0,784	-0,131/- 0,490	2,334/7,642	1,452/1,732
t-Testi p değeri	0,000	0,000	0,000	0,000	0,000	0,000
Mann-Whitney-U	0,399	0,000	0,000	0,000	0,000	0,000
Levene Homojenlik	0,086	0,121	0,003	0,006	0,065	0,072
K-S Normallik	0,000	0,064	0,062	0,334	0,045	0,005
Ki-kare Normallik	0,398	1,000	0,346	1,000	0,063	0,435

Moda ve haber siteleri olmak üzere iki bağımsız grup arasındaki anlamlı farklılığın incelenmesi için uygulanan t-testi sonucunda gruplar arası anlamlı farklılık olduğu görülmüştür. Elde edilen test sonucunun tutarlılığının kontrolü amacıyla gerçekleştirilen ve anakitle örnekleme dağılımına ait normallik şartının aranmadığı Mann-Whitney U testi aynı sonucu vermiştir. Normal dağılım durumunun incelenmesi için Ki-kare uygunluk ve Kolmogorov-Smirnov testleri uygulanmıştır. Testler sonucunda, eğiklik basıklık katsayılarının normal dağılımdan sapma gösterdiği görülmüştür. Farklı kategorilerine göre incelenen değişkenlere ait dağılım grafikleri Ağustos 2018 ve Ocak 2019 tarihleri için **Ek'te** verilmiştir. Benzetim çalışmalarında birden çok ilişkili değişken ile model kurma ihtiyacı olmaktadır. Göstergeler arasındaki ilişkiler incelenerek **Tablo.2'**de verilmiştir. Değişkenler arası ilişkinin doğrusal olmama durumuna karşı Spearman ve Pearson korelasyon katsayıları birlikte incelenmiştir.

**Tablo.2:** Kategorilere Göre Değişkenlere Ait Korelasyon Sonuçları

Korelasyon	Hemen Çıkma Oranı-Sayfa Sayısı		Hemen Çıkma Oranı-Sayfada Kalma Süresi		Sayfa Sayısı-Sayfada Kalma Süresi	
	Haber	E-Ticaret	Haber	E-Ticaret	Haber	E-Ticaret
Pearson	-0,783	-0,602	-0,779	-0,412	0,689	0,798
Spearman	-0,913	-0,678	-0,889	-0,560	0,819	0,871

Haber sitelerinden hemen çıkma oranının, sitede kalma süresi ve sayfa görüntüleme sayısı değişkenleri ile ters yönlü yüksek derece ilişki içerisinde olduğu görülmüştür. Aynı zamanda sayfa görüntüleme sayısı ile sitede kalma süresi pozitif yönlü yüksek ilişki içerisindedir. E-ticaret sitelerinde hemen çıkma oranının, sayfa

görüntüleme sayısı ve sitede kalma süresi değişkenleri ile ters yönlü orta derece ilişki içerisinde olduğu görülmüştür. Spearman korelasyon katsayısının her değişkende daha güçlü sonuçlar vermesi nedeniyle, ilgili değişkenler arasındaki ilişkinin doğrusal olmadığını söylemek mümkündür.

Dağılım incelenmesi amacıyla kullanılan ilgili yazılım, aynı anda 61 farklı teorik dağılım için parametreleri tahmin ederek uygunluklarını hesaplamaktadır. Parametrelerin belirlenmesinde ise en küçük kareler yöntemi, en çok benzerlik yöntemi, olasılık-ağırlıklı momentler ve L-momentler yöntemlerinden birini standart olarak kullanmaktadır (Seçkin, 2016). Uygunluğu belirlenerek parametreleri tahmin edilen dağılım özellikleri **Tablo.3**'te verilmiştir.

**Tablo.3:** Değişkenlerin Kategorilere Göre İstatistiksel Dağılım Sonuçları

Kategori Değişken	E-Ticaret Siteleri		Haber Siteleri		İncelenen Dağılımlar	
	Ağustos 2018	Ocak 2019	Ağustos 2018	Ocak 2019	Dağılım	Johnson SB
Sitede Kalma Süresi	<b>Dağılım</b> k= 0,80507 $\alpha$ = 3,4082 $\beta$ = 0,21975	<b>Dağılım</b> k=0,29954 $\alpha$ =4,2096 $\beta$ =0,29371	<b>Johnson SB Dağılımı</b> $\alpha$ =0.04165 $\beta$ =0.97989 $\gamma$ =64.239 $\delta$ =24.335	<b>Ters Gauss Dağılımı (3P)</b> $\alpha$ =0,98638 $\beta$ =0,26655 $\gamma$ =-0,07814	k: sürekli biçim parametresi ( $k>0$ ) $\alpha$ : sürekli biçim parametresi ( $\alpha>0$ ) $\beta$ : sürekli ölçek parametresi ( $\beta>0$ ) $\gamma$ : sürekli konum parametresi ( $\gamma=0$ ) $\delta$ : sürekli konum parametresi ( $\delta>0$ ) $\lambda$ : sürekli konum parametresi ( $\lambda>0$ ) $\xi$ : sürekli konum parametresi	
Hemen Çıkma Oranı	<b>Burr Dağılımı</b> k= 0,80234 $\alpha$ = 6,5053 $\beta$ = 29,559	<b>Dağılım</b> k=4,4936 $\alpha$ =2,9956 $\beta$ =20,36	<b>Johnson SB Dağılımı</b> $\alpha$ =0.04165 $\beta$ =0.97989 $\gamma$ =64.239 $\delta$ =24.335	<b>Johnson SB Dağılımı</b> $\alpha$ =-0.31689 $\beta$ =1.5719 $\gamma$ =121.18 $\delta$ =-2.1521		
Sayfa Görüntüleme Sayısı	<b>Dağılım</b> k= 0,53738 $\alpha$ = 5,2908 $\beta$ = 5,9679	<b>Dağılım</b> k=0,44352 $\alpha$ =4,2354 $\beta$ =5,6958	<b>Dağılım</b> k= 0,53738 $\alpha$ = 5,2908 $\beta$ = 5,9679	<b>Log-Pearson (3P) Dağılımı</b> $\alpha$ =22.813 $\beta$ =0.1101 $\gamma$ =-1.747		

Elde edilen tahminlerin güvenilir ve doğru şekilde gerçekleştirilebilmesi adına teorik dağılım modelinin seçimi ve seçilen modelin parametrelerinin uygun bir şekilde belirlenmesi gerekmektedir. Bu çalışmada ise internet sitelerinden elde edilen serilerde uygulanan Dağılım modellerinden hangisinin daha uygun olduğunu belirlemek amacıyla uygunluk testlerinden Ki-Kare testi (Yıldırım, 2013), Kolmogorov-Smirnov (Smirnov, 1939) ve Anderson Darling (Seçkin, 2016) testleri uygulanmıştır.

Dağılım dağılımı, genelleştirilmiş lojistik-Burr dağılımı modelini ifade etmesi nedeniyle, **Ters Burr dağılımı** veya **Kappa dağılımı** da denilmektedir. Aslında,  $\beta=1$  olması durumunda, Dağılım dağılımı log-lojistik dağılım olarak adlandırılmaktadır (Rajasooriya, 2013). Dağılım dağılımı, özellikle gelir dağılımının modellenmesinde log-normal ve gamma dağılım sınırlarına alternatif olarak, finans ve ekonomi alanlarında ve stokastik baskınlık çalışmalarında kullanılmaktadır. Johnson SB dağılımı, dört parametrelili sürekli bir dağılım olmakla birlikte, özellikle günümüzde 3 parametrelili Weibull dağılımı ile birlikte, model yapılarının oldukça esnek olmaları sebebiyle farklı modellemelerdeki başarılarıyla diğer dağılım fonksiyonlarına göre öne çıkmaktadırlar. Bu dağılım, epidemiyolojik araştırmalar ve risk değerlendirmesi, ormancılık alanında çap dağılımı belirlemede, yönelem alanında Markov zincirinin simülasyonunun yapılmasında öbekleme yönteminde ilk geçiş zamanlarının belirlenmesinde, hava

kirliliği ve yağış dağılımlarında vb. pek çok alanda ön plana çıkan bir dağılım türüdür. Bu dağılım, aslında Johnson SU dağılımından türetilmiş olan özel sınırlı dağılım olmakla birlikte, normal dağılımın basıklığının daha az olduğu durumlarda oluşan Platykurtik dağılımlara uygunluk göstermektedir.

**Tablo.3'te** parametreleri verilen dağılımlardan Dagum'a ait olasılık yoğunluk fonksiyonu aşağıda verilmiştir. "3P" üç parametrelili, "2P" iki parametrelili ve "4P" dört parametrelili dağılımı ifade etmektedir. Bu değişim durumu, örneğin bu çalışmada kullanılan değişkenlerin negatif değere sahip olmaması nedeniyle, dağılımın alt ve üst sınırlarının  $-\infty$ 'a yaklaşmaması için, dağılım parametre sayısını değiştirerek uyarlanmaktadır.

$$f(x) = \frac{\alpha k \left(\frac{x}{\beta}\right)^{\alpha k - 1}}{\beta \left(1 + \left(\frac{x}{\beta}\right)^\alpha\right)^{k+1}} \rightarrow 3P \qquad f(x) = \frac{\alpha k \left(\frac{x-\gamma}{\beta}\right)^{\alpha k - 1}}{\beta \left(1 + \left(\frac{x-\gamma}{\beta}\right)^\alpha\right)^{k+1}} \rightarrow 4P$$

Johnson SB dağılımına ait olasılık fonksiyonu da aşağıda verilmiştir.

$$\left(\xi \leq x \leq \xi + \lambda\right)$$

$$f(x) = \frac{\delta}{\lambda \sqrt{2\pi z(1-z)}} \exp\left(-\frac{1}{2}\left(\gamma + \delta \ln\left(\frac{z}{1-z}\right)\right)^2\right)$$

Elde edilen sonuçlar doğrultusunda, e-ticaret ve haber sitelerine ait eski ve yeni tarihli değişkenlere ait dağılımlar olasılık yoğunluk fonksiyonları elde edilerek grafikleri sunulmuştur. İlgili web sitelerine ait değişkenlerin, eski ve yeni dağılımları arasındaki anlamlı bir fark olup olmaması durumunun incelenmesinde Minimumlaştırılmış Kullback-Leibler Uzaklık Oranı'na dayalı ayırım analizi uygulanmıştır. İlk olarak 1951 yılında Kullback ve Leibler tarafından tanımlanan ve istatistikte bir X tesadüfi değişkeni için  $f_1(x)$  ve  $f_2(x)$  gibi iki dağılıma ait yoğunluk fonksiyonları arası uzaklığa ait simetrik olmayan bir ölçüdür. Bromideh ve Valizadeh (2013) tarafından, herhangi bir veri setinin, hangi muhtemel dağılımdan modellenebileceğinin belirlenmesi adına Minimumlaştırılmış Kullback-Leibler Uzaklıkları Oranı (RMKLD) yöntemi önerilmiştir. Bu yöntem ile dağılımlara ait doğru seçim olasılıkları belirlenmektedir. Yönteme ait hipotez, test istatistiği ve karar aşaması aşağıda verilmiştir (Bromideh ve Validezah, 2013).

$$H_0: X \sim f_1(x, \cdot); H_a: X \sim f_2(x, \cdot) \qquad RMKDL = \ln \left[ \frac{D_{KL}(f_1||f_2)}{D_{KL}(f_2||f_2)} \right]$$

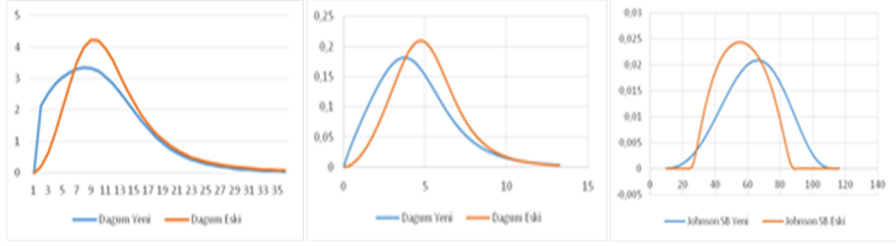
Test istatistiği  $RMKDL < 0$  ise  $H_0$  hipotezi red edilemez. RMKDL yöntemi için eski ( $f_2$ ) ve yeni ( $f_1$ ) tarihli dağılımlara ait dağılımların olasılık yoğunluk fonksiyonlarından log-benzerlik değerleri elde edilmiştir. Log-benzerlik değerlerinin elde edilmesinde gözlenen dağılımın ne kadar muhtemel olduğunu ifade eden Gauss formülünden yararlanılmıştır.

$$L = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\bar{y})^2}{2\sigma^2}} \propto e^{-\chi^2} \text{ ve genel olarak}$$

$$L_{distr.}(\theta_1, \theta_2, \dots, \theta_k; x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_k) \text{ şeklinde yazılır.}$$

Ki-kare değerinin en aza indirgenmesi, log-benzerlik olasılığının en üst seviyeye çıkarılmasıyla eşdeğerdir. Formül normal dağılım olması durumunda tam sonucu verirken, farklı dağılımlarda kullanılması durumunda mümkün en yakın sonuca

ulaşılmasını sağlamaktadır. Zaman içinde aynı dağılıma sahip olan değişkenlerin dağılımları arasındaki benzerlik ve farklılıkların incelenmesi amacıyla oluşturulan grafikler **Şekil-1**'de sırasıyla verilmiştir.



Şekil-1-a) Siteye Kalma Süresi Değişkeni O.y.f. b) Sayfa Görüntüleme Sayısı O.y.f. c) Hemen Çıkma Oranı O.y.f.

**Şekil-1-a)**'da görüldüğü üzere, e-ticaret sitelerinde kalma süresindeki homojenlik durumu Ocak 2019'da Ağustos 2018'e göre bozularak daha fazla farklılık göstermiştir. Ağırlığın daha çok düşük sürelerle doğru kaydığı görülmüştür.  $RMKDL=-2,84<0$  olduğundan e-ticaret sitelerinde kalma süresi dağılımında anlamlı bir değişiklik gözlenmemiştir. **Şekil-1 b)**'de görüldüğü üzere, e-ticaret sitelerinde sayfa görüntüleme sayısındaki homojenlik durumu Ocak 2019'da Ağustos 2018'e göre bozularak daha fazla farklılık göstermiştir. Ağırlığın daha çok düşük sürelerle doğru kaydığı görülmüştür. Kullanıcıların sayfalar arası gezinme ihtiyacının yükseldiği ancak buna rağmen mod sayfa sayısında düşüş olduğu görülmüştür.  $RMKDL=5,52>0$  olması nedeniyle temel hipotez ret edilmektedir. E-ticaret sitelerinde sayfa görüntüleme sayısının dağılımında anlamlı bir değişiklik ortaya çıkmıştır. **Şekil-1-c)**'de görüldüğü üzere, haber sitelerinden hemen çıkma oranındaki homojenlik durumu Ocak 2019'da Ağustos 2018'e göre bozularak daha fazla farklılık göstermiştir. Bununla yanı sıra dağılım parametrelerinden olan " $\xi$ " değerinin eski tarihe göre yükselmesi sonucu hemen çıkma oranlarında da yükseliş görülmektedir. Test istatistiği  $RMKDL=2,41>0$  olması nedeniyle temel hipotez red edilmektedir. Haber sitelerinden hemen çıkma oranının dağılımında anlamlı bir değişiklik gözlenmektedir.

## 5. Sonuç

Çalışma kısıtlarından bir tanesi, incelenmesi hedeflenen web sitelerine ait veriye ulaşma imkânlarının yeterli olmaması nedeniyle, incelenen haber ve moda web sitelerinin sayıca fazla olmayışıdır. Bundan dolayı, araştırma kapsamında incelenen web sitesi sayısının artması sonucu, araştırma sonuçlarının da farklılık gösterebileceği ihtimali göz ardı edilmemelidir. Web sayfalarının hit ve tıklanma miktarları da reklamlara ve sosyal mevcudiyet ve sosyal etki teorisine bağlıdır. Çalışmadaki bir diğer kısıt ise, bahsedilen durumdan ötürü, incelenen web sitelerine ait kişi başına giriş yapılmaması elde edilememesi nedeniyle bu veriye ait dağılım özelliklerinin incelenememesidir. Sonuç olarak, Türkçe yayınlanan web sitelerinden çıkarım yapılması nedeniyle toplumsal ve kültürel farklılıklar içerebilmekte ve ülkeler arası karşılaştırmanın yapıldığı, toplumun elektronik ortamdaki alışveriş ve haber takip davranışları değerlendirilebilmektedir.

Çalışmada kullanılan program ile fizik, coğrafya, tıp vb. farklı disiplinlerde yer alan araştırmacılar tarafından kullanılmasının yanı sıra istatistik alanında da farklı dağılımların kullanılabilirliği gösterilmiştir. İncelenen değişkenlerin ilişkili olmasıyla

marjinal dağılımlar elde edilmiş olup buradan hareketle de, yapılacak başka çalışmalarda koşullu dağılımlarında elde edilebileceği mümkün görülmektedir. Örneğin elde edilebilecek koşullu bir dağılım ile web sitelerinde ki kullanıcıların siteden hemen çıkma oranlarının 1 olması durumunda okunma sayısının kaç olacağı bilgisine ulaşılabilecektir. Bunun yanı sıra yöntem olarak bakıldığında, bayesyen dağılım yöntemlerinin de kullanılması mümkündür. Değişkenlere ait elde edilen dağılım bilgileriyle, özellikle özel sektörde kullanılan satış simülasyonlarında, multimedya satış sorumlularının kendini performanslarını değerlendirmesinde, web siteleri arasındaki performans değerlendirmesi gibi pek çok alanda kullanılabilceği düşünülmektedir.

Elde edilen bulgular doğrultusunda, kullanıcıların internet kullanımının daha amaç odaklı artış gösterdiği düşünülmektedir. Hedef odaklı kullanım gerçekleştirildiğinin düşünülmesi, görüntülenen sayfa sayısının azaldığının ve bununla birlikte hemen çıkma oranının da azalmış olduğunun görülmesiyle desteklenmektedir. Bu nedenle hemen çıkma oranı ve sayfa görüntüleme sayısı göstergelerinin, yapılacak benzetim modellerinde Dagum ve Johnson SB dağılımları ile incelenebilir olduğu çıkarımı yapılabilmektedir. Ancak sitede kalma süresi göstergesinde anlamlı bir farklılaşma henüz söz konusu olmadığından bu yorumun yapılması mümkün gözükmemektedir. İnternet kullanıcılarının davranış şeklinin değişmesi nedeniyle, araştırmacıların diğer web sitesi göstergeleri için sürekli güncel verilere dayalı dağılımları kullanarak değerlendirmeler yapmaları uygun olacaktır.

Sonuç olarak, web sitesi odaklı istatistiklerin incelendiği bu çalışmada, sayfa merkezli verilere ait dağılımların elde edilmesi sonucu planlama ve benzetim modellerine kaynak oluşturulabilmesi amacı gerçekleştirilmiştir.

## KAYNAKÇA

- ADEYEMI O.J. at al. (2018), *Exploration of Daily Internet Data Traffic Generated in a Smart University Campus, Data in Brief*, DOI: 10.1016/j.dib.2018.07.039i
- BARNDORFF-NIELSEN, O. E. (1997). *Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling*. Journal of Statistics. 24: 1-13.
- BUCKLIN, R. E., SISMEIRO, C. (2009). *Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing*, Journal of Interactive Marketing. 23, 35-48.
- BROMIDEH, A. A. ve VALIZADEH, R. (2013). *Discrimination Between Gamma and Log-Normal Distributions by Ratio of Minimized Kullback-Leibler Divergence*. Pakistan Journal of Statistics and Operation Research. 9.4.
- ÇİLİNGİRTÜRK, A. M. ve BOZKURT, G. (2018). *İnternet Sitesi İstatistiklerinin Dağılımı*. Marmara Üniversitesi 19. Uluslararası Ekonometri, Yöneylem Araştırması ve İstatistik Sempozyumu. 1171-1184.
- DALESSANDRO, B., HOOK, R., PERLICH, C. ve PROVOST, F. J. (2012). *Evaluating and Optimizing Online Advertising: Forget the Click, but There are Good Proxies*.
- HEUER, D., BRETTEL, M., KEMPER, J. (2015), *Brand Competition in Fashion E-Commerce. Electronic Commerce Research and Applications*. 14, 514-522.
- HUBERMAN, B.A., PIROLI, P., PITKOW, J.E. ve LUKOSE, R. (1998). *Strong Regularities in World Wide Web Surfing. Science*. 280.3, 95-97.
- JOHNSON, E.J., BELLMAN, S. ve LOHSE, G.L. (2003). *Cognitive Lock-In and the Power Law of Practice*. Journal of Marketing. 67.2, 62-75.
- LEVINA, O., VILNAI-YAVETZ, I. (2015), *E-Visibility Maturity Model: A Tool for Assessment and Comparison of Individual Firms and Sets of Firms in E-Business*. Electronic Commerce Research and Applications. 14, 480-498.



NOTTORF, F. (2014), *Modelling the Clickstream Across Multiple Online Advertising Channels Using a Binary Logit with Bayesian Mixture of Normals*. Electronic Commerce Research and Applications. 13, 45-55.

RAJASOORIYA, S. (2013). *The Mc-Dagum Distribution and Its Statistical Properties With Applications*. Asian Journal of Mathematics and Applications. Volume 2013. 1-16.

SEÇKİN, N. (2016). Bursa İlinde Gözlenen Yıllık Maksimum Yağışların Bölgesel Frekans Analizi. İMUCO – Nisan 21-22, 69-80.

SMIRNOV, N.V. (1939). *On the Estimation of the Discrepancy Between Empirical Curves of Distribution for Two Independent Samples*. Bulletin of Moscow University. 2.2, 3-14.

ÜLKEN, İ. (2010). *Doğu Karadeniz Havzası'nda En Uygun Dağılımla Debi Süreklilik Eğrilerinin Bulunması*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.

YILDIRIM, N. (2013). *Normal Dağılım İçin Uyum İyiliği Testleri ve Bir Simülasyon Çalışması*. Yayınlanmamış Yüksek Lisans Tezi, Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü.

### **Summary**

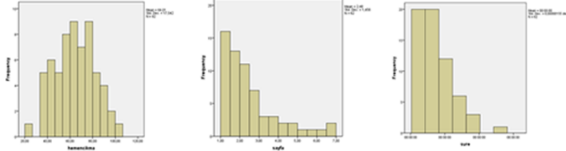
*The concept of internet, which has come into our lives with developing technology, has changed the communication channels in marketing as in many other disciplines. A larger communication network has emerged with changing communication channels between institutions and customers. As a result of rapid and high development, quality network services need to be provided and planned. There is a need for accurate estimation of Internet-based data. As a result of this need, experimental and empirical models can be produced based on real life situations.*

*The aim of this study is to model the distributions of page-based data consisting of the click data, and to create a source for simulation and planning models. For this purpose, the target web sites, accessories and ready-to-wear e-commerce and information is intended to obtain information in the form of news sites. When we look at the related literature, there are two different opinions that these web sites of different structure differ due to content differentiation in terms of time on the website and in terms of the number of pages visited by users, or in contrast, not differing. The web sites of the website were compared with the web sites. The main objective of the study is to determine whether the distribution of website statistics has changed in the short term and to propose to focus on certain distributions in planning.*

*According to the findings, users' use of the Internet is thought to be more objective-oriented. Considering goal-oriented use is supported by the fact that the number of pages displayed decreases and the bounce rate is also reduced. For this reason, it can be concluded that the bounce rate and page view count indicators can be examined by Dagum and Johnson SB distributions in the simulation models to be made. However, it is not possible to make this comment since there is no meaningful variation in the duration of stay on the site. Due to the change in the behavior of Internet users, it would be appropriate for the researchers to make evaluations using other distributions based on current data for other website indicators. As a result, in this study, which is focused on website focused statistics, it is aimed to create a resource for planning and simulation models as a result of the distribution of page based data.*

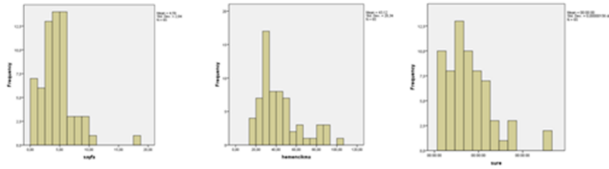
## EKLER

### EK-1: Değişkenlerin Haber Siteleri Kategorisine Ait Histogram Grafikleri-Ocak 2019



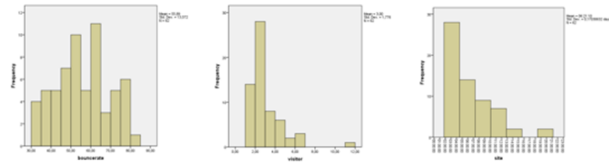
Histogramları verilen değişkenlerden, hemen çıkma oranı değişkeni 64,01 ortalama ve 17,54 standart sapma, sitede kalma süresi değişkeni 08:23 ortalama ve 06:13 standart sapma ile sayfa görüntüleme sayısı değişkeni ise 2,47 ortalama ve 1,45 standart sapmayla dağılım göstermektedir.

### EK-2: Değişkenlerin E-Ticaret Siteleri Kategorisine Ait Histogram Grafikleri-Ocak 2019



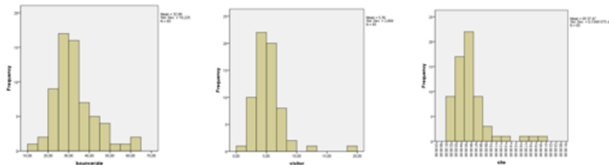
Histogramları verilen değişkenlerden, hemen çıkma oranı değişkeni 43,12 ortalama ve 20,34 standart sapma, sitede kalma süresi değişkeni 06:22 ortalama ve 04:37 standart sapma ile sayfa görüntüleme sayısı 4,59 ortalama ve 2,83 standart sapmayla dağılım göstermektedir.

### EK-3: Değişkenlerin Haber Siteleri Kategorisine Ait Histogram Grafikleri-Ağustos 2018



Histogramları verilen değişkenlerden, hemen çıkma oranı değişkeni 55,89 ortalama ve 13,572 standart sapma, sitede kalma süresi değişkeni 3,00 ortalama ve 1,776 standart sapma ile sayfa görüntüleme sayısı 06:21:10 ortalama ve 0,170 standart sapmayla dağılım göstermektedir.

### EK-4: Değişkenlerin E-Ticaret Siteleri Kategorisine Ait Histogram Grafikleri-Ağustos 2018



Histogramları verilen değişkenlerden, hemen çıkma oranı değişkeni 43,80 ortalama ve 16,568 standart sapma, sitede kalma süresi değişkeni 05:37:56 ortalama ve 0,154 standart sapma ile sayfa görüntüleme sayısı değişkeni ise 5,30 ortalama ve 2,669 standart sapmayla dağılım göstermektedir.