

Benford Yasasının Veri Doğruluğunun Değerlendirilmesi Amaçlı Kullanımı: Hastane Verileri İçin Bir Uygulama

Mustafa Said YILDIZ*

ÖZ

Bu çalışmanın amacı literatürde daha sıklıkla hile denetiminde kullanılan Benford Yasası'nın veri doğrulaması için kullanımını bir uygulama ile göstermektir. Sağlık planlaması, kaynak tahsisi ve performans değerlendirmesinde kullanılan hastane yönetsel verilerinden üçünün (poliklinik sayısı, acil servise başvuru sayısı ve röntgen çekim sayısı) Benford Yasasına uygunluğu ki-kare analizi ile belirlenmiştir. Uygulaması kolay bu analizin veri doğruluğunun değerlendirilmesi amacıyla, daha yüksek kalitede veri için kullanılmasının fayda sağlayacağı tespit edilmiştir.

Anahtar Kelimeler: Benford Kanunu, Veri Doğrulama, Hastane Verisi

JEL Sınıflandırması: C60, P46

Examining Data Accuracy With Benford's Law: An Application for Hospital Data

ABSTRACT

The purpose of this study is to demonstrate Benford Law's -which has been used frequently used for fraud auditing- for data validation. Conformity of three hospital administrative data (number of polyclinic visits, emergency department visits and x-ray transactions) which is used for healthcare planning, resource allocation and performance evaluation with Benford's Law is determined by chi-square test. It is determined that, utilization of this conveniently practicable tool for data validation would be beneficial for higher quality data.

Keywords: Benford's Law, Data Validation, Hospital Data

JEL Classification: C60, P46

GİRİŞ

Faaliyetlere ilişkin verilerin sağlıklı ve doğru olması gerek geçmişe dönük değerlendirmeler, gerekse planlamalar için büyük önem taşımaktadır. Veri işleme, toplama ve depolamanın teknolojik gelişmelerle giderek kolaylaşmasına ve sayısal verilerin nicelik ve niteliğinde artışların yaşanmasına rağmen verilerin sağlıklı olup olmadığının değerlendirilmesine hala ihtiyaç bulunmaktadır (Martin ve Conte, 2012). Ayrıca bilimsel çalışma verilerinin sağlıklı şekilde toplanıp toplanmadıkları (Diekmann, 2007:323), fabrikasyon verinin kullanılıp kullanılmadığı (Bassam, 2003) gibi hususlarda doğrulama prosedürleri gereksinimi bulunmaktadır. İlk olarak 1881 yılında ortaya konulan ve 1938 yılında kullanımı yaygınlaşmaya başlayan Benford Yasası, kullanım kolaylığı ve denetim alanında genel kabul görmüş olması nedeniyle veri güvenliği değerlendirmesi yapmak isteyenlere elverişli bir araç sunmaktadır.

* Dr. İç Denetçi, Sağlık Bakanlığı, msaid.yildiz@saglik.gov.tr

(Makale Gönderim Tarihi: 06.09.2017 / Yayına Kabul Tarihi: 11.12.2018)

Doi Number: 10.18657/yonveek.336919

Bu çalışmada Benford Yasasının veri doğrulaması alanında kullanımına bir örnek olmak üzere; kamu, özel ve üniversite hastanelerinden Sağlık Bakanlığına gönderilen üç veri setinin (poliklinik sayıları, acil başvuru sayıları ve röntgen çekim sayıları) Benford dağılımına uygunluğu değerlendirilecektir. Türkçe literatürde Benford Yasasının henüz veri doğrulama için kullanılmadığı göz önünde bulundurulduğunda çalışmanın katkı sağlayacağı düşünülmektedir. Uygulamaya rehberlik edebilmesi için olabildiğince yalın ve basit bir anlatım şekli tercih edilmiş, bilgi kaynakları ve erişim yollarına da yer verilmeye gayret edilmiştir.

I. BENFORD YASASI TARİHİ, TANIMI ve LİTERATÜRDE YERİ

Bu bölümde öncelikle Benford Yasası'nın ortaya çıkışı ve matematiksel altyapısı üzerinde durulacak, farklı veri setlerinde yapılan denemelerine örnekler verilecektir. Uluslararası ve Türkçe literatürde Benford Yasasının kullanım alanlarına, artan popülerliğine ve veri doğrulamasında kullanımlarına da ayrıca değinilecektir.

A. Benford Yasası'nın Tarihi ve Matematiksel Altyapısı

Astronom ve fizikçi Simon Newcomb (1881) American Journal of Mathematics'de kaleme aldığı makalesinde ilk olarak ortaya atılan düşünce yıllar içinde farklı çalışmalarla geliştirilmiştir. Newcomb daha sonra yasa olarak benimsenecek olan durumu ilk olarak sezgisel olarak tespit ettiğini ifade etmektedir. Hesaplamalar için logaritma tablolarının kullanıldığı dönemde sayfaların yıpranma derecesinin ilk sayfalarda daha fazla olduğu, yıpranmanın sonraki sayfalara doğru giderek azaldığı gözlemi ona ilham vermiştir. Kitabın yalnızca ilk sayfalarını okuyan da, sonraki sayfalara ulaşabilenler de, kitabı tamamlayanlar da ilk sayfaları yıpratacak, sonraki sayfalar giderek daha az oranda yıpranacaktır. Makalesinde yıpranmanın eşit sıklıkta olması beklenirken; en fazla 1'li basamakların yıprandığı ve yıpranmanın 9'lu basamaklara doğru azaldığına yer vermiştir. Buna göre çok sayılı setlerde ilk basamağı 1 olan sayılar, 2 olanlardan, 2 olan sayılar 3 olanlardan ve 3 olan sayılar 4 olanlardan... nihayetinde ilk basamağı 8 olanlar 9 olanlardan daha sıklıkla bulunacaktır.

Newcomb'un bu savı sonraki yıllarda da çeşitli çalışmalara konu olsa da 1938 yılında General Electrics'de fizikçi olarak çalışan Frank Benford (1938) tarafından farklı veri kümelerinden toplam 20.229 sayıyı derlediği çalışmasında ciddi şekilde yeniden gündeme gelmiştir. Bu makalede Benford nehir uzunlukları, nüfus, gazete tirajları, ölüm oranı... gibi pek çok veri için ilk basamakların frekanslarına yer vermiş; söz konusu frekans değerlerinin 1'den 9'a kadar düzenli şekilde azaldığını göstermiştir. Söz konusu olan nehirlerin uzunluğu, nüfus veya ölüm oranı... hangi veri seti olursa olsun; ilk basamağı 1 olanlardan 9 olanlara kadar giderek azalma göstermekteydi. Yani verilerdeki rakamların 1 ile başlama ihtimali 9 ile başlama ihtimalinin neredeyse altı katı idi ve tüm rakamların yüzde 30'unun üzerinde bir büyüklük oluşturuyordu. Halbuki bu rakamların 1,2,3,...9 ile başlama olasılıklarının oldukça yaklaşık olması beklenirdi. Zira Bassam

(2003) yaptığı çalışmada bilgisayar tarafından üretilen 10.000 rakama, yine bilgisayar tarafından üretilen 700'den fazla rakama ve insan eliyle üretilen rakamlara yer vermiş, bu rakamların ilk basamaklarının söz konusu 9 sayıya yaklaşık olarak denk sayıda dağıldığı görülmüştür (Bassam, 2003: 123). Doğal olarak toplanmış rakamlar Benford oranlarıyla dağılırken üretilmiş rakamların eşit dağıldığı böylelikle görülmüştür.

Bassam çalışmasında rakamların ilk basamaklarının 1'den 9'a kadar dağılımları Tablo 1'de gösterilmiştir. Benford oranlarının dağılışı ile Bassam çalışması sonucu üretilmiş veri rakamlarının yaklaşık olarak yakın orandaki dağılımları arasındaki farklılık görülebilecektir.

Tablo 1: Benford Yasasına uygun dağılım ile üretilmiş rakamların dağılımı arasında farklılık

	İlk basamaklar								
	1	2	3	4	5	6	7	8	9
Benford rakamları	30.10	17.60	12.50	9.70	7.90	6.70	5.80	5.10	4.60
Bilgisayar tarafından üretilen rakamlar (n=10.000)	11.60	11.40	11.30	11.00	11.20	11.10	11.00	10.40	11.00
Bilgisayar tarafından üretilen rakamlar (n=742)	10.90	11.70	12.00	12.30	11.50	11.20	10.80	11.10	8.6
Kaynak: Benford (1938), Bassam (2003)									

Benford Kanunu'nda yer alan oranların nasıl belirlendiğine değinmek faydalı olacaktır. Her bir rakamın yüzdesine o rakamın basit bir logaritma fonksiyonuna yerleştirilmesi ile ulaşılmaktadır. Benford Kanunu basitçe sayıların ilk basamaklarının eşit şekilde dağılmak yerine logaritmik formüle göre dağıldıklarını ifade eder. Buna göre ilk basamak D_1 olmak üzere;

$$d = 1, 2, 3, 4, 5, \dots, 9 \text{ için; } Pr(D_1 = d) = \log_{10}[1 + 1/d]$$

Örneğin, ilk basamağın 3 olması olasılığının 3 ise $\log_{10}[1 + 1/3]$ yaklaşık 0.1249 olması beklenecektir. Bu durumda logaritma formülünde "d" yerinde 1, 2, 3... 9 rakamlarını kullanmak suretiyle birinci basamağı 1, 2, 3... 9 olma olasılıkları belirlenebilecektir. Formülde d'nin paydada yer alması nedeniyle rakamlar 1'den 9'a doğru büyüdükçe, ilgili basamakta yer alma olasılığı azalacaktır (Lesperance vd. 2016).

Bir veri kümesinin Benford Kanunu'na göre dağılım göstermesinin şartları yasanın uygulamada yaygınlaşması için önemli pay sahibi olan Nigrini tarafından aşağıdaki gibi sıralanmıştır (Nigrini, 2000).

- Veri kümesi benzer olguların büyüklüğünü tanımlamalıdır.
- Veri kümesinin alt ve üst limitleri olmamalıdır. Örneğin mesai saatlerinde çalışma süresi istatistikleri günde en fazla 8 saati aşamayacağından bu veride Benford uygulanamaz.
- Değerler önceden belirlenmiş, bir kurala göre atanmış rakamlar olmamalıdır. Örneğin vergi numarası, telefon numarası vb.

B. Benford Yasasının Kullanımına İlişkin Literatür

Farklı veri setlerine ilişkin basamak sayılarının Benford Yasası'na uygunluğu görülmekle yasaya ve kullanımına güven artmıştır. Aşağıdaki tabloda farklı veri setlerinin basamaklara Benford yasasına uygun şekilde dağılımları

görülebilmektedir (Testing Benford Law web sitesi, 2017). Çok farklı alanlardan alınan bu rakamların Benford yasasının öngördüğü durumu yansıttığı görülebilmektedir.

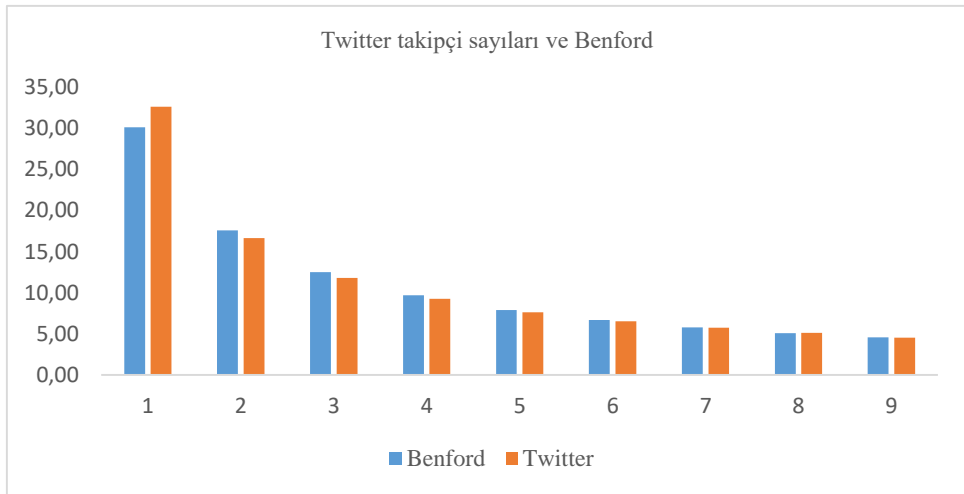
Tablo 2: Farklı veri setlerinin ilk basamaklarının Benford Yasasına uygun dağılımları

	İlk basamaklar (%)								
	1	2	3	4	5	6	7	8	9
Twitter kullanıcı sayıları	32.62	16.66	11.8	9.26	7.63	6.55	5.76	5.14	4.56
Türkiye'nin ilçe nüfusları	29.16	20.09	12.31	11.56	8.75	4.75	5.29	4.43	3.67
ABD kütüphaneleri yayın sayısı	33.73	17.86	11.07	8.65	7.67	5.98	5.25	5.29	4.50
Farklı ülkelerin nüfusları	27.41	16.29	12.30	10.61	9.34	6.84	6.53	5.38	5.37
İlk 652066 Fibonacci sayısı	30.1	17.61	12.49	9.69	7.92	6.69	5.80	5.12	4.58

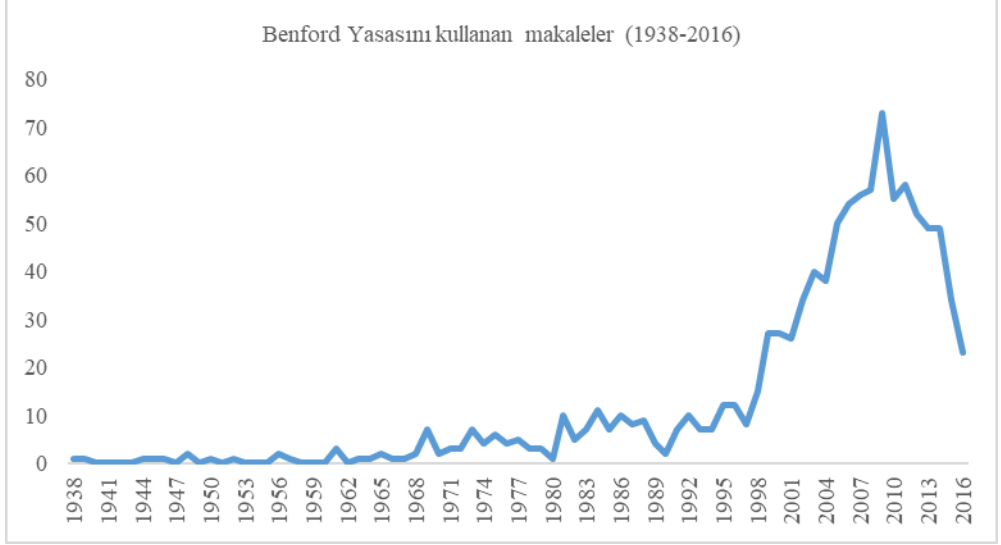
Kaynak: <http://www.testingbenfordslaw.com>

Grafik 1'de twitter sosyal medya platformunu takipçi sayılarının Benford Yasasında öngörülen rakamlarla karşılaştırılmasına yer verilmiştir. Takipçi sayılarının her bir rakamla (1, 2, 3...9) başlama yüzdelerinin Benford'da öngörülene çok yakın olduğu görülebilmektedir (Testing Benford Law web sitesi, 2017).

Grafik 1: Twitter takipçi sayılarının basamaklara dağılımının Benford Yasasına Uygunluğu



Newcomb tarafından 1881 yılında kullanılması sonrasında 1938 yılına kadar beş ayrı makaleye konu olabilen bu matematiksel gerçeğin 1938 yılı sonrasında yıllar itibariyle bilimsel yayınlara konu olma sayılarına Grafik.2'de yer verilmiştir (Benford Online Bibliography, 2017). Grafikten yasanın kullanımının giderek 2000'li yıllardan başlayarak artış gösterdiği anlaşılmaktadır.

Grafik 2: 1938-2016 yılları arasında Benford Yasası kullanılarak yazılan makaleler

Kullanım kolaylığı ve kullanılabilmesi verilerin çeşitliliği nedeniyle Benford Yasası çok çeşitli alanlarda kullanılmıştır. Doğal afet zararları veri tabanının homojenliğinden (Joannes-Boyau vd. 2015) online sosyal ağlarda arkadaş ve takipçi sayılarına (Golbeck, 2015), finansal raporların denetiminden (Henselmann vd. 2013) seçim hileleri (Deckert vd, 2011) ve belirsizliklerinin (Breunig ve Goerres,2011) belirlenmesine bir çok alanda akademik makalelere konu olan Benford giderek popülerliğini artırmıştır. ABD'nin önemli gazetelerinden Wall Street Journal'ın adli muhasebecilerin Benford'u sıklıkla kullandıklarına ilişkin Aralık 2014 tarihli haberi (Wall Street Journal web sitesi, 2014), Princeton Üniversitesi Yayınları tarafından 2015 yılında yayınladığı kitaplar (Miller, 2015) (Berger ve Hill, 2015). Mozy ve arkadaşlarının (2016) saygın akademik dergilerden Nature'da 2016'da yayınlanan bilimsel rapor ve bibliyografyalar ışığında yukarıda yer verilen grafik söz konusu popülerliğin devam ettiğine işaret etmektedir.

Uluslararası literatürde mali suç ve yolsuzlukla mücadelede sıklıkla kullanılmıştır. Toshiba yolsuzluğu'nun tespitinde yasanın kullanıldığı çalışma (Mehta ve Bhavani, 2017), bir finansal suç için veri çalışması (Carreira ve Gomes da Silva, 2016), Busta ve Weinberg (1998) tarafından sinir ağları ve Benford yasası kullanarak etkin bir değerlendirme prosedürü oluşturmaya çalıştıkları çalışma ve mali piyasaların Benford'a göre değerlendirildiği Corazza ve arkadaşlarının (2010) çalışması bunlara örnek olarak alınabilir.

Benford yasasının sağlık alanında da çeşitli amaçlarla kullanıldığı gözlemlenmektedir. MR ve BT görüntülerinin yeniden yapılandırılması (Sanches ve Marques, 2006), ölüm oranlarının gösterge olarak kullanımının güvenilirliğinin

sorgulanması (Pollach vd. 2015) ve sağlık hukukçuları tarafından klinik veri yolsuzluğunun tespiti (Health Lawyers web sitesi, 2017) amaçlı kullanımlar bunlardan bazılarıdır.

Benford yasasının kullanım amaçlarından biri de veri doğrulamasıdır. Çeşitli amaçlarla toplanan verinin usulüne uygun şekilde mi toplandığı yoksa herhangi bir gerçekliğe dayalı olmaksızın mı üretildiği hususu verinin Benford yasasına uygun olup olmadığı test edilerek belirlenmeye çalışılmıştır.

Judge ve Schechter (2009) çalışmalarında anket verisindeki problemler ve anormallikleri değerlendirmişler ve sıklıkla kullanılan dokuz veri setinde yaptıkları değerlendirmede gelişmekte olan ülke verilerinin daha düşük, ABD’de toplanan verilerin daha yüksek kalitede oldukları; erkek ve bayan cevaplayanlar arasında ise farklılık bulunmadığı sonucuna ulaşmışlardır. Anket sonuçlarının kalitesi ve güvenilirliğinin istatistiksel olarak da değerlendirilebileceği göz önünde bulundurulduğunda, istatistiksel değerlendirme sonucunda daha yüksek kalitede bulunduğu görülen sonuçların Benford analizini teyid etmesi söz konusu olmuştur. Benford yasasına uygun olmama durumu ise doğrudan verinin fabrikasyonla üretildiği suçlaması yapmak için yeterli olmamakla beraber bu duruma karine oluşturan analizlerden biri olarak öne çıkabilecektir. Nitekim anket sonucu etik dışı müdahaleler sonucu üretilmemiş olsa dahi örneklemin uygun şekilde belirlenmemiş oluşu, tesadüfi seçilmemiş olması gibi çeşitli sebeplerle düşük kalitede veri oluşabilecektir.

Diekmann (2007) ise makalesinde bilimsel çalışmada kullanılan verilerin uydurulmuş – “fabrikasyon” olması ihtimalini değerlendirmiş. Bunun için de verilerin Benford yasasının öngördüğü dağılımlara uygun olup olmadığını test etmiştir.

Martin ve Conte (2012) mali planlamaya esas verinin kalitesini değerlendirmek üzere Benford yasasından yararlanmışlardır. Aslına uygun ve özgün olup olmadığı (authenticity) Bassam (2003) tarafından da Benford yasası kullanılarak araştırılmıştır.

Benford’un hesaplanması ve kullanımının izah edildiği birçok kaynağa erişilebilir. ISACA (Uluslararası Sistem Denetimi ve Kontrolü Derneği - Information Systems Audit and Control Association)’nın web sitesi IIA (İç Denetçiler Enstitüsü - The Institute of Internal Auditors) web sitesi Journal of Accountancy’de Collins (2017) tarafından kaleme alınan yazı Benford’un kullanımının izah edildiği kaynaklardan bazılarıdır. Benford Kanunu yardımıyla veri validasyonu için kullanılmak üzere yazılım paketleri de üretilmiştir. Açık kaynak R yazılımı için CRAN (The Comprehensive R Archive Network) tarafından üretilen kodlar bunlardan birisidir (CRAN Benford yazılımı, 2017).

Türkiye’de bir makaleye ilk olarak 2001 yılında konu olan Benford yasası (Erdoğan, 2001) sonrasında pek çok konuda makalelerde ve tezlerde kullanılmıştır. Türkçe literatürün incelenmesinde, Benford yasasının muhasebe ve

vergi denetimi ile hilenin ortaya çıkarılması alanlarında kullanımına vurgu yapıldığı ve bu alanlardan örneklerle uygulamalar yapıldığı görülmektedir. Akkaş(2007) denetim alanında Benford uygulamasına ilişkin çalışmasında literatüre değinilmiş ve bir uygulama ile Benford'un kullanımı örneklendirilmiştir. Alagöz ve Ay (2002) ile Boztepe (2013) çalışmalarında yine muhasebe denetiminde Benford Kanunu'nun uygulamasına yer verilmiştir. Yanık ve Samancı (2013) kamu muhasebe verileri için Benford'u kullanmıştır. Avcı ve Demirci (2016) ise vergi denetiminde bir uygulaması ile literatüre katkı yapmıştır. Cinko (2014) Borsa İstanbul'un günlük getirileri için yasayı uygulamış ve sonuçlarını sunmuştur. Benford Yasası yıllar içinde çeşitli alanlardan yüksek lisans tezlerine de konu olmuştur. İlk tezler 2004 (Çakır, 2004) ve 2006 (Kocameşe, 2006) yıllarında yapılan tezleri, 2014 yılında 3 (Samancı, 2014) (Demir, 2014) (Özdemir, 2014) ve 2016 yılında bir tez (Rasgen, 2016) takip etmiştir. Söz konusu tezlerin de tümünün denetimde Benford'un kullanımına ilişkin olduğu görülmüştür.

Benford Yasasının veri kalitesinin değerlendirilmesi amaçlı kullanımına ilişkin literatür veya uygulama çalışmasına ise rastlanılmamıştır. Bu yönüyle bu çalışmanın Türkçe literatüre Benford Yasasının farklı bir amaçla kullanımına ilişkin bir uygulama sunarak katkı sağlayacağı umulmaktadır.

II) VERİ DOĞRULAMA ALANINDA BİR UYGULAMA

Çalışmanın bu bölümünde; 2012 yılında Sağlık Bakanlığı Sağlık Hizmetleri Genel Müdürlüğüne ulaşan veriler sayısal analiz testlerinden en geneli olan, birinci basamak testine dayalı bir analiz yapıldıktan sonra ki kare uygunluk testine tabi tutulmuştur. Bu testlerin sonucunda ortaya çıkan veriler ile Benford değerleri karşılaştırılmış ve verilerin sağlıklı şekilde iletilip iletilmediği, verilerde hata olup olmadığı incelenmiştir. Böylelikle veri validasyonu için Benford Yasasından nasıl yararlanılabileceğine ilişkin bir uygulama gerçekleştirilecektir.

Hastanelerin yönetsel verileri planlama, kaynak tahsisi, performans değerlendirme gibi pek çok amaçlarla kullanılmaktadır. Bu verilerden 1079 adet kamu, özel ve üniversite hastanesinden gelen üçünün (yıllık poliklinik sayıları, acil servis hasta başvuru sayıları, kapalı alan ve röntgen çekim sayısı) sağlıklı şekilde iletilip iletilmediği bu bölümde Benford yasası ile sınanarak değerlendirilecektir.

Uygulama için 1079 hastanenin verileri birinci basamağına göre gruplandırılmış; her bir rakamla (1,2,3,4,...9) başlayan veri adedi belirlenmiştir. Poliklinik sayılarının ilk rakamlara dağılımının Benford dağılımına uygun olduğu grafikte görülebilmektedir. Acil servise başvuru ve röntgen çekim rakamları için aynı yöntem izlendiğinde yine Benford yasasına uyumlu bir dağılımın gerçekleştiği görülebilmektedir.

Her bir basamakta beklenen Benford sayılarının aşağıdaki formülle belirlendiğini tekrar etmekte fayda vardır.

$$d = 1, 2, 3, 4, 5 \dots 9 \text{ için; } Pr (D_1 = d) = \log_{10}[1 + 1/d]$$

Örneğin Benford Yasasına göre 2 ile başlaması beklenen sayıların oranı için $\log_{10}(1+1/2)$ 'in 100 ile çarpımı 17,61 değeri belirlenmiştir. Diğer Benford değerleri de aynı şekilde 3 ile başlayan için $\log_{10}(1+1/3)$, 4 ile başlayan için $\log_{10}(1+1/4)$, ... 9 ile başlayan için $\log_{10}(1+1/9)$ şeklinde belirlenmiştir. Gözlemlenen poliklinik, acil ve röntgen çekim sayıları ise her bir sayı ile (1,2,3...9) başlayan verilerin toplam veri içindeki yüzdesidir.

Tablo 3: Hastanelerin yıllık poliklinik başvurusu, acil başvuru ve röntgen çekim sayılarının basamaklara yüzde olarak dağılımı ve Benford'dan farklılıkları

	Benford-(B) (%)	Poliklinik Başvurusu-PB (%)	Fark B- PB (%)	Acil Başvuru- AB (%)	Fark (B- AB) (%)	Röntgen Çekim Sayısı- RC (%)	Fark (B- RC) (%)
1	30,10	35,03	-4,93	26,51	3,59	31,21	-1,11
2	17,61	14,27	3,34	17,67	-0,07	18,38	-0,77
3	12,49	9,36	3,13	13,77	-1,27	11,70	0,79
4	9,69	8,90	0,79	10,88	-1,19	7,19	2,50
5	7,92	7,14	0,78	8,37	-0,45	7,39	0,53
6	6,69	7,32	-0,63	7,35	-0,65	7,60	-0,90
7	5,80	6,67	-0,87	6,14	-0,34	6,16	-0,36
8	5,12	6,77	-1,65	5,49	-0,37	6,37	-1,25
9	4,58	4,54	0,03	3,81	0,76	4,00	0,57

Benford Yasası'nın öngördüğü yüzde dağılım ile yıllık poliklinik sayıları acil başvuru sayıları ve röntgen çekim sayılarının büyük ölçüde uygunluk taşıdığı görülebilmektedir. En fazla uygunluk (Benford yasasından en az oranda sapma) röntgen çekim sayısında görülmüş; acil başvuru oranlarında yalnızca ilk basamağı 1 olanlar, poliklinik başvuruları oranlarında ise ilk basamağı 1, 2 ve 3 olanlar daha fazla sapma göstermişlerdir. Bu noktada farkların mutlak değer toplamlarına bakmak sapmaların boyutunu görmek için faydalı olacaktır. **Poliklinik başvurusu verisinde 16,16 düzeyinde mutlak değer farkı bulunurken; acil başvuru sayısı ve röntgen çekim sayısı verilerinde birbirine oldukça yakın olarak sırasıyla 8,71 ve 8,81 mutlak değer farkı bulunmuştur.**

Gözlenen (poliklinik başvurusu, acil başvuru ve röntgen çekim sayısı) dağılım ile beklenen (Benford yasasına uygun) dağılım arasında farkların büyüklüğüne ki-kare test istatistiği (uyum iyiliği - goodness of fit) ile bakılabilir (Yıldız ve Akbulut, 2014). Ki-kare değeri bulunduktan sonra tablo değeri ile karşılaştırma yapılabilecektir. Uygulamaya rehberlik edebilmesi ve kullanım pratikliğinin artırılması için analiz istatistik yazılımı kullanılmadan gerçekleştirilmiştir.

H_0 : Sağlık Bakanlığına ulaşan hastane verileri Benford Kanunu'na uygun dağılmaktadır.

H_1 : Sağlık Bakanlığına ulaşan hastane verileri Benford Kanunu'na uygun dağılmamaktadır.

O: Gözlenen Değer, E: Beklenen Değer

$$\chi^2_{\text{test}} = \sum \frac{(O_i - E_i)^2}{E_i}$$

Tablo 3: Poliklinik başvuru sayısına ait ki-kare değerinin ki-kare formülü kullanılarak hesaplanması

	Beklenen (Benford) E _i	Gözlenen Poliklinik Başvurusu O _{ip}	O _i -E _{ip}	(O _i -E _{ip}) ²	(O _i -E _{ip}) ² /E _i
1	30,1	35,03	4,93	24,3049	0,807472
2	17,61	14,27	-3,34	11,1556	0,633481
3	12,49	9,36	-3,13	9,7969	0,78438
4	9,69	8,9	-0,79	0,6241	0,064407
5	7,92	7,14	-0,78	0,6084	0,076818
6	6,69	7,32	0,63	0,3969	0,059327
7	5,8	6,67	0,87	0,7569	0,1305
8	5,12	6,77	1,65	2,7225	0,531738
9	4,58	4,54	-0,04	0,0016	0,000349
Ki-kare değeri					3,088472
Gözlem sayısı: 1079					

Böylelikle gözlenen poliklinik sayıları için ki-kare değeri 3,088472 hesaplanmıştır. Aynı şekilde gözlenen acil başvuru ve gözlenen çekim sayısına da uyguladığımızda acil başvuru için ki-kare değeri 0,972501, röntgen çekim sayısı için ki-kare değeri ise 1,32978 olarak hesaplanmaktadır. İlk basamağı 9 rakama dağılan veriler için 9 adet grubun ki-kare uygunluk testinin serbestlik derecesi 8'dir. Ki-kare cetvelinde kritik değer sekiz serbestlik derecesinde %1 anlamlılık seviyesinde 20,09 ve %5 anlamlılık seviyesinde ise 15,51 değeridir. **Ki-kare tablosundaki kritik değerden oldukça düşük bulunan her üç veri için H₀ hipotezi kabul edilecek; kamu, özel ve üniversite hastanelerinden (1079 adedi için) Sağlık Bakanlığına bildirilen verilerin Benford Yasasına uyumlu olduğu sonucuna ulaşılmıştır. Sonucun Benford yasasına uyumlu olmaması durumunda doğrudan bir yargılama ile verinin hatalı olduğu sonucunda şüphesiz varılamayacak idi. Bu durumda dağılıştan uzaklaşılabilir parçalarda bir yeniden değerlendirme yapılarak bu durumun sebebi üzerinde durmak gerekebilecekti. Bu durumda Benford yasasının verinin sağlıklı olduğuna bir delil olarak değerlendirilmesi gerektiği, sağlıklı olduğuna ilişkin ise yalnız karinelere ve işaretlerden yalnızca biri olarak değerlendirmenin sağlıklı olacağı ifade edilebilir.**

SONUÇ

Hastanelere ait yönetsel veriler sağlık yönetim ve politikasının planlama, kaynak tahsisi, performans değerlendirme gibi pek çok amacı için sıklıkla kullanılmaktadır. Bu verilerin girişlerinden, aktarımından veya kaydından kaynaklanan hatalar söz konusu işlevlerin gerçekleştirilememesine veya yanlış kararlar alınmasına sebep olabilmektedir.

Muhasebe ve vergi alanında hile denetiminde kullanımı oldukça yaygınlaşmış bulunan Benford Yasası'nın literatürde bir diğer kullanım alanı verilerin sağlıklı olup olmadığının kontrolüdür. Benford Yasası, genel kabul görmüş olması yanında kullanımının kolaylığı ve pratikliği nedeniyle veri toplayan, üreten ve işleyenlerin verilerinin fabrikasyon ve sahte olup olmadığını belirlemeleri için önemli bir gereç olarak öne çıkmaktadır.

Bu çalışmada kamu, özel ve üniversite hastanelerinden (1079 hastane) Sağlık Bakanlığına gönderilen ve Bakanlık tarafından pek çok amaçla ve sıklıkla kullanılan üç önemli veri değerlendirilmiştir. Gerek Benford yasası dağılımlarının mutlak değerleriyle gerekse ki-kare analiziyle yapılan değerlendirmede verilerin Benford dağılımlarıyla uyumlu olduğu görülmüştür. Verilerin ve özellikle tereddüt yaşananlarının Benford Yasası'na uygunluğunun değerlendirilmesi ile yasaya uygunluk taşımayanların yeniden değerlendirilmesi yoluna gidilebilecektir. Benford Yasasına uygunluk değerlendirmesi gerek genel kabul görmüş olması gerekse kolaylığı ve uygulama pratikliği ile öne çıkmaktadır. Veri validasyonu ve değerlendirmesi için Benford Yasası'na uyum kontrolünün yapılmasının faydalı olacağı görülmüştür.

Benford Yasası'nın veri toplama ve işlemeyle ilgili olarak görev yapanlarla tanıştırılması verilerin doğruluğu ve güvenilirliğine katkıda bulunacaktır. Veri toplamak üzere hazırlanan yazılımlara Benford Yasası modülü'nün eklenmesi de bir seçenek olarak değerlendirilmelidir.

KAYNAKÇA

- Akkaş M.E. (2007). Denetimde Benford Kanunu'nun Uygulanması. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 9(1):191 - 206
- Alagöz A., Ay M. (2002). Muhasebe Denetiminde Benford Kanunu Temelli Dijital Analiz. *Sosyal Ekonomik Araştırmalar Dergisi*, 2(4): 59-76.
- Avcı O., Demirci Z. (2016). Benford Kanunu'nun Vergi Denetiminde Kullanımı ve Bir Örnek Uygulama. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*, 5(7): 2232-2246
- Bassam H. (2003). Examining data accuracy and authenticity with leading digit frequency analysis, *Industrial Management & Data Systems*, 103(2): 121-125. <https://doi.org/10.1108/02635570310463447>
- Benford F. (1938). The Law of Anomalous Numbers, *Proceedings of the American Philosophical Society*, 78(4), 551-572
- Benford Online Bibliography (2017), <http://www.benfordonline.net>, (Erişim tarihi: 01.08.2017)
- Berger A., Hill, T.P. (2015). *An Introduction to Benford's Law*. Princeton University Press: Princeton, NJ. ISSN/ISBN:9780691163062.
- Boztepe E. (2013). Benford Kanunu ve Muhasebe Denetiminde Kullanılabilirliği. *LAÜ Sosyal Bilimler Dergisi*, 4(1): 73-83
- Breunig C., Goerres A., (2011). Searching for electoral irregularities in an established democracy: Applying Benford's Law tests to Bundestag elections in Unified Germany, *Electoral Studies* doi:10.1016/j.electstud.2011.03.005
- Busta B., Weinberg R. (1998). Using Benford's law and neural networks as a review procedure, *Managerial Auditing Journal*, 13(6): 356-366.

- Carreira P., Gomes da Silva C (2016). Assessing the omission of records from a data set using Benford's law. *Journal of Financial Crime*, 23(4): 798-805 <https://doi.org/10.1108/JFC-10-2015-0060>
- Cinko M., (2014). BIST-100 Getirileri Dağılımının Benford Kanunu ile Testi. *Journal of Economics, Finance & Accounting-JEFA*, 1(3): 184-191
- Collins C., (2017). Using excel and Benford's Law to detect fraud, Journal of Accountancy web sitesi, <http://www.journalofaccountancy.com/issues/2017/apr/excel-and-benfords-law-to-detect-fraud.html>
- Corazza M., Ellero A., Zorzi A. (2010) Checking financial markets via Benford's law: the S&P 500 case. In: Corazza M., Pizzi C. (eds) *Mathematical and Statistical Methods for Actuarial Sciences and Finance*. Springer, Milano
- CRAN Benford yazılımı (2017). <https://cran.r-project.org/web/packages/benford.analysis/benford.analysis.pdf>
- Çakır S., (2004). *Muhasebe Hilelerinin Tespitinde İstatistiksel Yöntemler (Benford Yönteminin İrdelenmesi)*, Yayınlanmamış Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü, İzmir.
- Deckert J., Myagkov M., Ordeshook P. C., (2011). Benford's Law and the Detection of Election Fraud, *Political Analysis* 19:245–268 doi:10.1093/pan/mpm014
- Demir M. (2014). *Benford Yasası ve Hile Denetiminde Kullanılması*. Yayınlanmamış Yüksek Lisans Tezi, İstanbul Ticaret Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Diekmann A. (2007). Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data, *Journal of Applied Statistics*, 34(3), 321-329. doi:10.1080/02664760601004940
- Erdoğan M. (2001). Muhasebe Hilelerinin Ortaya Çıkarılmasında Benford Yasası. *Muhasebe ve Denetim Bakış*, 1(3): 1-8.
- Golbeck J. (2015). Benford's Law Applies to Online Social Networks. *PLoS ONE*, 10(8): e0135169. doi:10.1371/journal.pone.0135169
- Health Lawyers web sitesi (2017), https://www.healthlawyers.org/events/programs/materials/documents/fc13/206_godse_mettin_slides.pdf (Erişim Tarihi: 16.08.2017)
- Henselmann K. Scherr E, Ditter D (2013) : Applying Benford's Law to individual financial reports: An empirical investigation on the basis of SECXBRL filings, Working Papers in Accounting Valuation Auditing, No. 2012-1 [rev.]
- İIA (İç Denetçiler Enstitüsü) web sitesi, [http://www.theiia.org/intAuditor/media/files/Step-by-step-Instructions-for-Using-Benford's-Law\[1\].pdf](http://www.theiia.org/intAuditor/media/files/Step-by-step-Instructions-for-Using-Benford's-Law[1].pdf)
- ISACA (Uluslararası Sistem Denetimi ve Kontrolü Derneği) web sitesi, <https://www.isaca.org/Journal/archives/2010/Volume-1/Pages/Using-Spreadsheets-and-Benford-s-Law-to-Test-Accounting-Data1.aspx>
- Joannes-Boyau R., Bodin T, Scheffers M, May S.M. (2015). Using Benford's law to investigate Natural Hazard dataset homogeneity. *Nature Scientific Reports*, 5, 12046, doi: 10.1038/srep12046 (2015).
- Judge G., Schechter L. (2009). Detecting problems in survey data using Benford's Law, *The Journal of Human Resources*, 44(1): 1-24 doi: 10.3368/jhr.44.1.1
- Kocameşe M. (2006). *Benford Kanunu ve Vergi Denetiminde Kullanılabilirliğinin İncelenmesi*. Yayınlanmamış Yüksek Lisans Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Lesperance M., Reed W.J., Stephens M.A., Tsao C., Wilton B. (2016). Assessing Conformance with Benford's Law: Goodness-Of-Fit Tests and Simultaneous Confidence Intervals. *PLoS ONE*, 11(3):e0151235, doi:10.1371/journal.pone.0151235
- Martin J, Conte T., Data quality in financial planning - an empirical assessment based on Benford's Law (2012). ECIS 2012 Proceedings. 159. <http://aisel.aisnet.org/ecis2012/159>
- Martin, J., Conte, T. (2012). Data Quality in Financial Planning - An Empirical Assessment Based On Benford's Law. *ECIS 2012 Proceedings*. 159. Retrieved from <http://aisel.aisnet.org/ecis2012/159>

- Mehta A., Bhavani G. (2017), Application of forensic tools to detect fraud: the case of Toshiba, *Journal of Forensic and Investigative Accounting*, 9(1): 692-710
- Miller, S.J. (2015). *Benford's Law: Theory and Applications*. Princeton; Oxford, Princeton University Press, JSTOR, Retrieved from: www.jstor.org/stable/j.ctt1dr358t.
- Morzy M., Kajdanowicz T., Szymański B.K. (2016). Benford's Distribution in Complex Networks. *Nature Scientific Reports*, 6, 34917; doi: 10.1038/srep34917
- Newcomb S. (1881). Note on the Frequency of Use of the Different Digits in Natural Numbers, *American Journal of Mathematics*, 4(1), 39-40.
- Nigrini, M. J. (2000). *Digital Analysis Using Benford's Law Tests & Statistics for Auditors*, Second Edition, Global Audit Publications, Canada. p.11
- Özdemir M. (2014). *Muhasebe Denetiminde Benford Kanunu ve Ölçekten Bağımsızlık Yönteminin Sınanmasına Yönelik bir İnceleme*. Yayınlanmamış Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Pollach G., Jung K., Namboya F. and Pietruck C. (2015) Maternal Mortality Rate—A Reliable Indicator? *International Journal of Clinical Medicine*, 6, 342-346. <http://dx.doi.org/10.4236/ijcm.2015.65044>
- Rasgen M. (2016). *Hile Denetiminde Benford Yasasının Kullanılmasına İlişkin Bir Uygulama*. Yayınlanmamış Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü, İzmir.
- Samancı T.H (2014). *Denetimde Benford Yaklaşımı*. Yayınlanmamış Yüksek Lisans Tezi, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü, Erzurum.
- Sanches J., Marques J.S. (2006). Image reconstruction using the Benford Law, *International Conference on Image Processing*, Atlanta, GA, 2006, pp. 2029-2032. doi: 10.1109/ICIP.2006.312845
- Testing Benford Law web sitesi (2017), <http://www.testingbenfordslaw.com/> (Erişim Tarihi: 16.08.2017)
- Wall Street Journal web sitesi (2017) Accountants Increasingly Use Data Analysis to Catch Fraud (2014, 05 Aralık). Wall Street Journal, <https://www.wsj.com/articles/accountants-increasingly-use-data-analysis-to-catch-fraud-1417804886> (Erişim Tarihi 16.08.2017)
- Yanık R, Samancı T (2013) Benford Kanunu ve Muhasebe Verilerinde Uygulanmasına ait Kamu Sektöründe Bir Uygulama, *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 17(1): 335-348
- Yıldız N., Akbulut Ö. (2014), İstatistiğe Giriş: Uygulamalı Temel Bilgiler, Çözümlü ve Cevaplı Sorular (5.Baskı). Erzurum: Aktif Yayınları.

SUMMARY

Accuracy and authenticity of hospitals' administrative data is crucial for planning, resource allocation and performance evaluations in healthcare. Although technology of collecting and transferring data improved in recent years, data validation for accuracy is still a substantial requirement. Data collection and recording processes need a convenient tool for validation. Benford Law which has been used frequently for fraud auditing can be used for data validation and fraudulent – fabricated data determination.

Digits of data which is produced by natural process without any intervention are distributed in a manner described by Benford's Law. Instead of distributing equally to numbers from 1 to 10; percentages of distribution follow Benford's logistic formula and decrease from 1 to 9. By testing the data sets for their conformance to this distribution, fraudulent and fabricated data can be discriminated from accurate and authentic data.

In order to demonstrate applicability of Benford's Law for data validation beyond fraud auditing; three hospital administrative data (number of polyclinic visits, emergency department visits and x-ray transactions) - which is used for healthcare planning, resource allocation and performance evaluation - is tested. These administrative data sets obtained from 1079 public, private and university hospitals were compared with Benford's distribution. Leading digit frequency analysis is conducted for each of the data sets. Initially percentages of each of leading digits (1,2,3...9) for data sets were determined, compared with leading digits proposed by Benford's Law and differences were reported. Conformity of three hospital administrative data with Benford's Law is determined by chi-square test. A significant conformance with Benford's distribution could be determined by the study which means a satisfying level of validity.

This study suggests that Benford's Law and the leading digit frequency analysis in particular offers a useful procedure for evaluation of data accuracy. Introducing Benford's Law to data collection professionals would be beneficial for accuracy of data. Integrating modules to data collection softwares which can automatically run Benford's Law and alerts in case of fabricated or fraudulent data could be another option for more accurate data acquisition.