

İSTATİSTİĞİN SÖYLE(ME)DİĞİ: BENFORD YASASINA UYUM(SUZLUK)

Dr. Gülşah Seydaoğlu

Tıp Doktoru, Biyoistatistik Doktorası, Çukurova Üniversitesi Tıp Fakültesi
Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Giriş

Son yıllarda, hem bilim hem de ticaret dünyasında yaşanan uygun olmayan istatistiksel yöntemler ve hileli verilere dair olguların giderek sık gözleniyor olması, bağımsız denetçileri, hileli ve yanıltıcı işlemlerin ortaya konulması amacıyla analitik araçların ve metotlarının geliştirilmesine yönlendirmiştir. Bu amaçla kullanılan araçlardan biri olan Benford Kanunu, olası hataların, düzensizliklerin veya potansiyel hilelerin-yanıltmaların ortaya konmasında basit ve etkin bir araç olarak kullanılmaktadır. Benford Kanunu'na dayalı sayısal analiz testleri, hatalı veya hileli verilerin büyük tutarlı veriler arasında aranması gerektiği ön yargısını yıkarak, düşük ve yüksek tutarların aynı risk düzeyinde analiz edilmesini sağlamaktadır. Bu analitik denetim tekniği, veriler arasında doğal kabul edilmeyecek sapmaları bulup ortaya çıkarmaktadır. Bu çalışmanın amacı, pandemi sürecinde 57 ülkede yayınlanan COVIF-19 vaka ve ölüm verilerinin Benford Kanunu'na uyumunu değerlendirmektir.

Materyal Metod

Benford Yasası şunu ifade etmektedir: Rakamların tümü incelendiğinde, ilk haneleri 1 ile başlayanlar yaklaşık bütün rakamların yüzde 30'unu, 2 ile başlayanlar yüzde 18'ini, 3 ile başlayanlar yüzde 13'ünü oluşturur ve böyle devam ederek 9 la başlayanların oranı yüzde 5 ten az olur.

Birinci Basamak	1	2	3	4	5	6	7	8	9
Benford Yasasının Olasılığı	0,301	0,176	0,125	0,097	0,079	0,067	0,058	0,051	0,046

Bu dağılım, sayıların gerçek olması halinde gerçekleşmektedir. Ancak sayılar gerçek değil, uydurma veya hatalı ise Benford Yasası çalışmamaktadır. ABD'de vergi dairelerinin mükelleflerin vergi beyanlarında veya seçim verilerinin sahteciliği, anlamak için bu yöntemi kullanılmaktadır.

Bu çalışmada veriler, 11 Mart 2020 ile 3 Eylül 2020 tarihleri arasında John Hopkins University veri tabanında yayınlanmış olan tablolardan elde edil-

miştir. Verilere dair ayrıntılı grafik ve tablolara Endüstri Mühendisi ve veri görselleştirme uzmanı Zeki Berk tarafından geliştirilmiş olan veri tabanından onun izniyle alınmıştır.¹⁸⁸ Zeki Berk tarafından geliştirilmiş olan bu web sayfasında Benford Yasası'na göre uyumsuzluk yüzdesi, demokrasi indeksi ile ilişkisi gibi ayrıntılı verilere ve grafiklere de ulaşmak mümkündür.

Benford Kanunu'na dayalı olarak hazırlanan, sayısal analiz testlerinin başlıcaları; Birinci Basamak Testi - İkinci Basamak Testi - İlk İki Basamak Testi - İlk Üç Basamak Testi - Tekrarlı Sayılar Testi - Son İki Basamak Testidir. Birinci basamak testi sayısal analizin ana testidir. Bu test uygunluk testi olup, çok geneldir. Bu çalışmada da Birinci Basamak Testi kullanılmıştır.

Benford Yasası olasılıkları ile hesaplanmış olan beklenen frekanslar ile ülkelere ait COVID-19 vaka ve ölüm için bildirilmiş olan gözlenen frekanslar arasındaki farkın istatistiksel olarak anlamlı olup olamayacağına ya da bu farkın rassal kabul edilip edilmeyeceğine karar vermek adına uygunluk için Ki-Kare test istatistiği kullanılmıştır.

Vaka ve ölüm sayılarının birinci basamak frekansları (digitleri) baz alınarak Ki-kare değeri formülü ile hesaplanmıştır. Burada : örnek büyüklüğü; : i rakamının gözlenen frekansı; : Benford yasasının olasılığıdır.

Ki-kare uygunluk testinin $r-1=9-1=8$ serbestlik derecesi göz önüne alınarak anlamlılık düzeyi $\alpha=0.05$ ve $\alpha=0.01$ alındığında tablo kritik değerleri sırasıyla 15.51 ve 20.09 olarak elde edilmiştir.

Aşağıdaki tabloda Çin vaka verileri kullanılarak örnek hesaplama gösterilmiştir. (Tablo 1). Çin örneğinde vaka sayılarının birinci basamak frekansları için Ki-Kare İstatistik değeri olarak elde edilmiştir. Bu değer, hem $\alpha=0.05$ hem de $\alpha=0.01$ düzeyindeki kritik tablo değerlerinden küçük olduğu için gözlenen frekanslar ile Benford Yasası'nın olasılıkları arasında bir uyum olduğu ve anlamlı bir fark olmadığı söylenebilir ($p>0,05$ ve $p>0,01$).

Tablo 1: Benford Yasasına Uyumluluk Örnek Hesaplama (Çin Vaka Verileri İle)

Ülke	Birinci Basamak	Basamak Frekansı (vaka)	Benford Yasasının Olasılığı ()	Beklenen Frekans ()	Ki-Kare İstatistik değeri
Çin	1	63	0,301	65,317	0,1
	2	36	0,176	38,192	0,1
	3	31	0,125	27,1250	0,6
	4	20	0,097	21,049	0,1
	5	22	0,079	17,143	1,4
	6	10	0,067	14,539	1,4
	7	14	0,058	12,586	0,2
	8	10	0,051	11,067	0,1
	9	11	0,046	9,982	0,1
Toplam		217			4,0

¹⁸⁸ <https://bit.ly/Benford-Test>

Bulgular

Günlük vaka ve ölüm verilerine ulaşılabilen toplam 57 ülkenin birinci basamak dağılımları ve bunların uyumları Tablo 2’de sunulmuştur.

Tablo 2: Ülkelerin Benford Yasası’na Uyum Durumu

Country	Group	First Digit									Chi-Square Statistic	P
		1	2	3	4	5	6	7	8	9		
Argentina	Cases	53	27	16	13	14	14	20	12	8	13,80	NS
	Deaths	64	26	17	15	12	6	10	9	9	7,70	NS
Armenia	Cases	41	31	27	26	23	11	13	3	2	23,7**	P<0.01
	Deaths	62	14	17	8	11	11	10	11	2	20,9**	P<0.01
Australia	Cases	66	38	27	19	7	7	13	9	8	9,50	NS
	Deaths	52	14	8	6	9	6	2	3	3	23,1**	P<0.01
Austria	Cases	48	32	34	14	19	9	14	14	6	12,80	NS
	Deaths	53	32	8	2	4	4	3	2	3	38,3**	P<0.01
Azerbaijan	Cases	48	16	38	19	30	8	8	4	4	44,9**	P<0.01
	Deaths	32	26	17	8	10	13	13	10	6	10,60	NS
Belarus	Cases	56	10	11	10	11	5	10	25	30	120,3**	P<0.01
	Deaths	3	8	19	38	50	27	8	0	0	250,9**	P<0.01
Belgium	Cases	57	27	19	15	18	17	9	11	9	5,3	NS
	Deaths	50	26	16	21	16	8	12	6	4	7,5	NS
Bolivia	Cases	77	20	13	9	6	16	6	16	10	36,5**	P<0.01
	Deaths	38	18	15	20	21	23	10	9	2	32,7**	P<0.01
Bosnia and Herzegovina	Cases	38	46	28	23	9	7	2	4	5	30,8**	P<0.01
	Deaths	49	28	15	11	8	10	2	4	5	9,7	NS
Brazil	Cases	44	36	28	36	20	9	5	1	4	38,3**	P<0.01
	Deaths	79	9	8	12	12	19	10	10	12	46,6**	P<0.01
Bulgaria	Cases	68	39	26	13	7	5	7	5	5	19,4*	P<0.05
	Deaths	43	26	22	18	8	7	7	10	4	5,9	NS
Canada	Cases	69	29	32	20	11	17	10	5	6	10,8	NS
	Deaths	64	15	16	18	14	17	13	6	10	16,9*	P<0.05
Chile	Cases	55	37	34	26	15	5	2	2	5	28,7**	P<0.01
	Deaths	42	21	13	20	14	16	12	10	17	22,1**	P<0.01
China	Cases	63	36	31	20	22	10	14	10	11	3,9	NS
	Deaths	50	24	16	15	5	6	5	5	1	12,7	NS
Colombia	Cases	58	23	21	13	7	15	12	14	14	15,2*	P<0.05
	Deaths	51	46	26	10	5	6	6	4	8	24,9**	P<0.01
Czechia	Cases	54	38	24	16	21	13	9	4	6	8,2	NS
	Deaths	41	30	14	12	12	2	5	6	3	10,7	NS
Denmark	Cases	74	18	22	17	11	6	12	6	10	18,9*	P<0.05
	Deaths	54	10	10	9	2	3	4	5	7	29,7**	P<0.01
Dominican Republic	Cases	51	30	23	21	10	10	7	13	8	5,3	NS
	Deaths	60	25	10	10	8	13	13	16	7	21,2**	P<0.01
Ecuador	Cases	47	20	8	14	18	16	12	9	18	30,5**	P<0.01
	Deaths	41	28	24	25	12	12	5	5	4	12,8	NS
Egypt	Cases	92	19	17	15	5	10	10	3	12	44,5**	P<0.01

Country	Group	First Digit									Chi-Square Statistic	P
		1	2	3	4	5	6	7	8	9		
El Salvador	Deaths	56	29	16	14	10	10	10	20	7	17,8*	P<0.05
	Cases	40	34	21	22	6	11	9	10	10	9,9	NS
France	Deaths	42	8	7	8	3	10	16	13	12	46,8**	P<0.01
	Cases	39	25	32	25	16	12	9	10	4	16,0*	P<0.05
Germany	Deaths	47	35	19	8	15	4	9	8	7	10,2	NS
	Cases	58	30	32	24	17	21	8	6	5	13,0	NS
Greece	Deaths	64	28	18	15	12	9	8	6	6	6,0	NS
	Cases	64	52	16	7	12	9	12	3	7	28,3**	P<0.01
Guatemala	Deaths	45	31	13	8	12	4	1	0	0	28,8**	P<0.01
	Cases	51	24	30	13	7	11	8	12	12	12,6	NS
Hungary	Deaths	43	33	26	15	7	2	1	2	0	32,5**	P<0.01
	Cases	46	35	30	14	14	11	11	9	9	5,0	NS
India	Deaths	54	24	11	7	7	6	1	8	3	20,0*	NS
	Cases	54	24	20	11	21	28	11	13	6	29,4**	P<0.01
Iran	Deaths	42	18	25	20	13	10	11	13	16	19,0*	P<0.05
	Cases	63	112	9	1	4	0	1	3	5	236,6**	P<0.01
Iraq	Deaths	101	20	4	9	13	14	17	9	10	61,6**	P<0.01
	Cases	37	51	37	19	10	13	8	9	3	30,5**	P<0.01
Ireland	Deaths	40	22	12	13	6	17	25	17	14	53,5**	P<0.01
	Cases	47	28	26	20	14	14	13	9	11	4,1	NS
Israel	Deaths	48	21	14	14	4	4	1	6	6	15	NS
	Cases	84	37	16	13	15	7	9	4	5	24,9**	P<0.01
Italy	Deaths	57	16	15	13	13	7	7	9	8	10,1	NS
	Cases	48	47	33	22	19	9	5	6	8	17,3*	P<0.05
Japan	Deaths	47	21	19	25	23	21	19	11	9	24,5**	P<0.01
	Cases	50	33	33	21	29	13	13	11	13	14,9	NS
Korea, South	Deaths	65	35	11	8	9	11	8	4	10	19,4*	P<0.05
	Cases	54	38	39	29	20	11	7	9	6	16,2*	P<0.05
Mexico	Deaths	53	26	14	6	5	9	3	4	2	18,9*	P<0.05
	Cases	37	20	25	31	25	27	11	4	2	54,3**	P<0.01
Moldova	Deaths	33	25	18	18	13	21	22	13	5	34,5**	P<0.01
	Cases	63	36	33	11	11	8	6	3	4	20,1*	P<0.05
Morocco	Deaths	47	11	10	15	20	19	17	11	7	34,2**	P<0.01
	Cases	88	25	8	5	12	10	13	9	10	42,4**	P<0.01
Netherlands	Deaths	50	35	22	13	8	2	6	2	0	24,7**	P<0.01
	Cases	68	17	13	13	20	23	16	12	6	30,5**	P<0.01
North Macedonia	Deaths	56	27	18	11	10	12	4	11	6	7,9	NS
	Cases	96	29	11	10	5	4	8	8	5	56,1**	P<0.01
Pakistan	Deaths	36	25	20	18	17	15	6	8	6	8,41	NS
	Cases	46	34	18	24	13	15	10	5	9	7,61	NS
Panama	Deaths	51	23	21	12	8	16	13	10	6	8,42	NS
	Cases	75	18	10	9	8	8	16	10	20	50,1**	P<0.01
Peru	Deaths	46	37	16	14	15	12	9	12	1	11,5	NS
	Cases	23	17	46	27	17	11	8	9	12	60,2**	P<0.01
	Deaths	90	28	5	6	6	5	5	6	5	64,8**	P<0.01

Country	Group	First Digit									Chi-Square Statistic	P
		1	2	3	4	5	6	7	8	9		
Philippines	Cases	49	52	25	20	11	10	10	3	5	20,0*	P<0.05
	Deaths	66	27	11	10	10	14	10	15	11	18,9*	P<0.05
Poland	Cases	14	37	57	21	22	12	10	5	4	91,6**	P<0.01
	Deaths	82	31	10	8	6	10	8	5	10	35,7**	P<0.01
Portugal	Cases	36	52	41	16	14	7	8	4	3	43,2**	P<0.01
	Deaths	50	42	32	6	7	14	5	6	8	23,9**	P<0.01
Romania	Cases	74	34	35	16	4	7	7	4	3	30,6**	P<0.01
	Deaths	59	45	26	16	6	7	1	2	4	31,3**	P<0.01
Russia	Cases	24	8	5	27	37	31	13	28	9	157,1**	P<0.01
	Deaths	99	9	1	8	6	7	9	7	16	102,2**	P<0.01
Saudi Arabia	Cases	85	33	28	18	4	1	2	3	6	48,1**	P<0.01
	Deaths	19	18	54	29	16	6	6	4	9	96,8**	P<0.01
Serbia	Cases	34	47	33	13	10	7	12	8	11	26,2**	P<0.01
	Deaths	57	11	22	11	15	10	9	7	9	14,9	NS
South Africa	Cases	57	30	23	15	6	15	10	14	9	8,7	NS
	Deaths	59	26	13	14	9	5	11	9	5	10,5	NS
Spain	Cases	32	23	28	21	21	13	14	14	5	22,5**	P<0.01
	Deaths	34	30	19	16	14	10	13	6	7	6,3	NS
Sweden	Cases	44	27	32	24	10	20	22	6	3	31,1**	P<0.01
	Deaths	36	23	21	18	12	11	18	20	7	28,6**	P<0.01
Switzerland	Cases	75	31	27	8	15	10	8	7	7	14,7	NS
	Deaths	46	26	15	11	13	3	7	1	0	17,8*	NS
Turkey	Cases	93	15	11	12	4	0	1	4	35	165,7**	P<0.01
	Deaths	75	43	8	11	8	5	8	3	9	36,9**	P<0.01
Ukraine	Cases	39	24	15	30	17	11	9	19	8	29,9**	P<0.01
	Deaths	80	38	12	6	9	1	3	6	8	47,3**	P<0.01
United Kingdom	Cases	62	19	22	30	23	18	15	11	7	19,2*	P<0.05
	Deaths	61	27	24	14	13	13	10	8	9	2,75	NS
US	Cases	33	59	33	27	18	21	5	2	0	58,6**	P<0.01
	Deaths	79	26	10	11	16	13	9	10	13	23,8**	P<0.01

*p<0.05 **p<0.01

Analize göre hem vaka hem de ölüm verileri yasa ile uyumlu olan ülke sayısı sadece 10'dur. 20 ülkede ise vaka (n=12) ya da ölüm (n=8) verileri yasaya uyum göstermemektedir. Toplam 57 ülkeden 27 ülkede ise hem vaka hem de ölüm verileri yasaya uymamaktadır (Tablo 3). Bu ülkeler arasında Güney Kore ve Filipinlerde istatistiksel farklılık hem vaka hem de ölüm verilerinde p<0.05 düzeyinde iken, diğer ülkelerde en az birinde (vaka veya ölüm verileri) farklılık p<0.01 düzeyindedir. Dolayısıyla, elde edilen bulgular kabul edilen anlamlılık düzeyinde, bu ülkelerin COVID--19 vaka ve ölüm kayıtlarının frekansları ile Benford kuramsal olasılıkları arasındaki farkın rassal kabul edilemeyeceğine, bu ülkelerde elde edilen verilerin güvenilir olmadığına işaret etmektedir.

Tablo 3: Ülke Verilerinin Yasa İle Uyum Durumu

Yasa İle Uyumlu Ülkeler	Vaka veya Ölüm Verilerinden Biri Uyumsuz Olan Ülkeler		Vaka ve Ölüm Verileri Uyumsuz
	Vaka Verileri Uyumsuz	Ölüm Verileri Uyumsuz	
Switzerland	North Macedonia	El Salvador	Iran
Argentina	Panama	Austria	Turkey
Germany	Azerbaijan	Guatemala	Russia
South Africa	Bosnia and Herzegovina	Colombia	Belarus
Czechia	Ecuador	Australia	Poland
Pakistan	Netherlands	Dominican Republic	Peru
Belgium	Serbia	Japan	US
Hungary	Israel	Canada	Mexico
Ireland	Spain		Saudi Arabia
China	Bulgaria		Egypt
	United Kingdom		Portugal
	France		Morocco
			Brazil
			Bolivia
			Sweden
			Romania
			Iraq
			Ukraine
			India
			Chile
			Greece
			Armenia
			Moldova
			Philippines
			Denmark
			Italy
			Korea, South

Ki Kare değerine göre sıralıdır.

Sonuç

Yaptığımız analize göre Türkiye de dahil olmak üzere dünyanın pek çok ülkesinin resmi olarak açıkladığı COVID-19 vaka ve ölüm kayıtlarının frekansları ile Benford kuramsal olasılıkları arasındaki farkın rassal kabul edilemeyeceği sonucuna varılmıştır. Bu verilerin uyumsuzluk nedenlerinin ülkelerin iktidarlarının popülist kaygıları, sağlık politikaları, demokrasi göstergeleri, tanrı, kayıt güvenilirliği ve virüsün seyri gibi bir çok faktör ışığında

tartışılmasında yarar vardır. Unutulmamalıdır ki; bilim dünyasında istatistiksel olarak yanıltma mümkündür, ancak bu durumun kontrolü veya ortaya konması yine istatistik ile mümkündür.

Seçilmiş Kaynak

- Benford, Frank. (1938), "The Law of Anonialous Numbers", Proceedings of American Philosophical Society, Vol.78, pp. 551-572.

BİZDEN HABERLER

