

Bu site, hizmet sunmak ve trafiği analiz etmek için Google'a ait çerezleri kullanır. Hizmet kalitesinden emin olunması, kullanım istatistikleri oluşturulması, kötüye kullanımın algılanması ve çözüme kavuşturulması amacıyla performans ve güvenlik metriklerinin yanı sıra IP adresiniz ve kullanıcı aracınız Google ile paylaşılır.

DAHA FAZLA BİLGİ TAMAM

Blog Arşivi

► 2016 (3)

▼ 2015 (2)

▼ Aralık (2)

İstatistikte
Bilmemiz
Gereken
Terimler

İstatistik ve
Araçları

31 Aralık 2015 Perşembe

İstatistikte Bilmemiz Gereken Terimler

Veri

Veriler ölçüm, sayım, deney, gözlem ya da araştırma yolu ile elde edilmektedir. Ölçüm ya da sayım yolu ile toplanan ve sayısal bir değer bildiren veriler nicel veriler, sayısal bir değer bildirmeyen veriler de nitel veriler olarak adlandırılmaktadır.

Table 3-1 Number of Crimes, Estimated Population Size, and Crime Rates in the U.S.

Year	No. of Crimes	Population Size	Crime Rate per 100,000 People
1998	12,475,634	270,296,000	4,615.5
1999	11,634,378	272,690,813	4,266.5
2000	11,608,072	281,421,906	4,124.8
2001	11,876,669	285,317,559	4,162.6
2002	11,878,954	287,973,924	4,125.0
2003	11,826,538	290,690,788	4,068.4
2004	11,679,474	293,656,842	3,977.3
2005	11,565,499	296,507,061	3,900.6
2006	11,401,511	299,398,484	3,808.1
2007	11,251,828	301,621,157	3,730.5
2008	11,149,927	304,059,784	3,667.0

Source: U.S. Crime Victimization Survey

Nümerik Veriler

Veriler ölçüm, sayım, deney, gözlem ya da araştırma yolu ile elde edilmektedir. Ölçüm ya da sayım yolu ile toplanan ve sayısal bir değer bildiren veriler nicel veriler, sayısal bir değer

bildirmeyen veriler de nitel veriler olarak adlandırılmaktadır. Örnek olarak, yan taraftaki veri setinde; ofislere göre gelirler gösterilmiştir. Bu kısımda gelir nümerik bir veridir.

Office	Revenue	Revenue
Whalen	\$4,400.00	\$4,400.00
Hartstein	\$13,000.00	\$13,000.00
Fay	\$6,000.00	\$6,000.00
Raphaely	\$11,000.00	\$11,000.00
Khoo	\$3,100.00	\$3,100.00
Baida	\$2,900.00	\$2,900.00
Tobias	\$2,800.00	\$2,800.00
Himuro	\$2,600.00	\$2,600.00
Colmenares	\$2,500.00	\$2,500.00
Mavris	\$6,500.00	\$6,500.00

Nominal (Kategorik) Veriler

Kategorik bir veri çeşidir. "Daha fazla" ifadesi ile kullanılmazlar.

İkiye ayrılır:

a) Dikotom Veriler :
Var-Yok, Kadın-Erkek, Hasta-Sağlıklı

Office	Revenue	Revenue
Whalen	\$4,400.00	\$4,400.00
Hartstein	\$13,000.00	\$13,000.00
Fay	\$6,000.00	\$6,000.00
Raphaely	\$11,000.00	\$11,000.00
Khoo	\$3,100.00	\$3,100.00
Baida	\$2,900.00	\$2,900.00
Tobias	\$2,800.00	\$2,800.00
Himuro	\$2,600.00	\$2,600.00
Colmenares	\$2,500.00	\$2,500.00
Mavris	\$6,500.00	\$6,500.00

Bu site, hizmet sunmak ve trafiği analiz etmek için Google'a ait çerezleri kullanır. Hizmet kalitesinden emin olunması, kullanım istatistikleri oluşturulması, kötüye kullanımın algılanması ve çözüme kavuşturulması amacıyla performans ve güvenlik metriklerinin yanı sıra IP adresiniz ve kullanıcı aracınız Google ile paylaşılır.

DAHA FAZLA BİLGİ TAMAM

Oranmaktadır.

Ordinal Veriler

Ordinal Data			
Point	Airports	Oil well production	Populated places
	✗ international ✗ national ✗ regional	■ high ■ medium ■ low	● large ● medium ● small
Line	Roads	Drainage	Boundaries
	— expressway — major — local	— river — stream — creek	— international — provincial — county
Area	Soil quality	Cost of living	Industrial regions
	■ good ■ fair ■ poor	■ high ■ medium ■ low	■ major ■ minor

Ordinal veriler de yine kategorik veri türündendir. Fakat değerleri arasında sıralı bir ilişki bulunmaktadır. “Daha fazla” ifadesi ile kullanılabilirler ancak ne kadar daha fazla olduğunun ölçüsünü veremezler. Örneğin: Eğitim

Düzeyi, Sosyo ekonomik ölçek skorları gibi. Nominal veriler, ordinal verilere daha az bilgi taşırlar.

Interval (aralıklı) veriler

Nesnelerin sıralanmasında kullanılır. Eşit aralıkların eşit mesafelerini temsil ettiği bir ölçek türüdür. Ordinal ölçümün bütün özelliklerini taşır. Nesneler arasındaki farkın mukayesesine imkan tanır. Başlangıç noktası sorunu vardır. Araştırmacı kendi yaptığı çalışmaya göre bir başlangıç noktası belirler. Sıfır noktası sabit olmadığı için başlangıç noktası görecelidir. Sıfır noktası sayısallaştırmayı kolaylaştıran bir başlangıç noktasıdır. Örneklere bakalım; tansiyon ölçümleri (küçük ve büyük tansiyon belirlenerek kişinin tansiyonunun bu değerler arasında olacağı açıklanır), kan pH'ı 7.1 ile 7.5 arasında olması, vücut sıcaklığını ölçmek için kullanılan termometreler, likert ölçeği (tutum ve davranışları ölçmek için 1 ile 5 arasında değer vermek gibi, ya da kötü—harika tarzı anlamsal dizilimler).

Ratio (Oransal) Veriler

Interval,ordinal ve nominal ölçüm türünün özelliklerini taşır. Aynı zamanda mutlak sıfır noktasına da sahiptir. En üst ölçüm tekniğidir, her türlü istatistiksel ve matematik işleme imkan tanımaktadır. Cevaplayıcı sıfırdan herhangi bir sayıya kadar cevap verebilir. Oransallık söz konusudur. Örnek vermemiz gerekirse; yıllık kazancınız ne kadar, kaç çocuğunuz var, sağlık alanında ağırlık ölçümü, boy ölçümü, fizik bilimindeki ölçümler ağırlık, alan, hacim gibi, ölen hastalanan yaralanan hakkındaki veriler vb.

DEĞİŞKEN (Variable)

- Bir değişken karakteristik veya nümerik değer olabilir bu deneyden deneye değişir.
- Bir değişken size bir sayımı (Örneğin; sahip olduğun evcil hayvanların sayısı) veya bir ölçümü yansıtabilir (Örneğin; sabahları uyanma saatlerinizin ölçümü).
- Veya değişkenler her bireyin belirli ölçütlere göre gruplandırıldığı kategorik verilerden oluşabilirler (Örneğin; politik görüş, ırk ya da medeni durum).
- Gözlem bilgilerinin, üzerine kaydedilen gerçek veri parçalarıdır.

DAHA FAZLA BİLGİ TAMAM

tamamını içmektense ilk önce çorbayı karıştırır ardından tencerenin içinden bir kaşık yardımıyla çorbadan



biraz alarak onu tadarsınız. Ve sonra bütün bir çorbayla alakalı bir yorum yapabilirsiniz. Eğer dürüst bir yolla alınmışsa, yani çorbanın sadece iyi yerleri alınmamışsa, çorba hakkında iyi bir fikir oluşturabilirsiniz. İşte İstatistikte raslantısal örnek seçimide aynı bu şekilde olmaktadır. Bir popülasyonu çorba olarak kabul edersek o çorbayı çok iyi karıştırıp öyle örnek seçmemiz gerekir ki doğru sonuca daha yakın olabilme şansımız olsun.

Parametre (Parameter)

İstatistik örnek verilere dayanır, popülasyon verilerine değil. Eğer verimizi tüm popülasyondan toplamışsak, bu toplama işlemine sayım denir. Eğer biz bu sayımı sayıyla bir değişken üzerinden özetliyorsak , bu sayı istatistik değil parametre olur.

Ortalama (Average or Mean)

Ortalama istatistikçilerce nümerik verilerdeki , bir ölçümün merkezi veya oratasını bulmak için kullanılmaktadır. Aritmetik ortalama bütün değerlerin toplamının, değer sayısına bölünmesiyle bulunmaktadır. Ortalama her zaman gerçeği yansıtmayabilir, çünkü ortalama aykırı değerlerden kolayca etkilenebilmektedir. (Aykırı değer : Çok küçük veya çok büyük değerler)

Medyan (Median)

Medyan veri setinin merkezini ölçmek için kullanılan diğer bir yoldur. Basitçe bir veri grubu küçükten büyüğe doğru sıralandığında, ortada bulunan veri ortanca değer (medyan) olarak adlandırılır.

Standart Sapma (Standard deviation)

Bir istatistiksel çalışma yaparken beklediğimiz kesin sonuç normalin üstünde veya altında bulunabilir. Zamanla insanlar bu durumu standartlaştırmak istemişler ve bunun için standart sapmayı bulmuşlar. Bu sayede Standart Sapma , yaptığımız çalışmanın sonuçlarının ortalamasının ne kadar üstüne ve altına sapabileceğini hesaplamamıza olanak sağlamıştır. Formülize edersek;

$$s = \sqrt{\sum \frac{(x - \bar{x})^2}{n-1}}$$

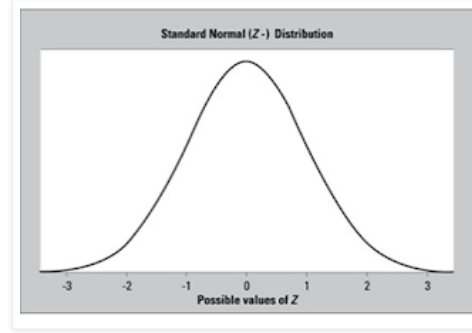
Diagram illustrating the formula for Standard Deviation (s) with labels:

- Gözlem** (Observation) points to x .
- Ortalama** (Average) points to \bar{x} .
- Gözlem Sayısı** (Number of Observations) points to n .

normal distribution)

Dağılım bir veri setinin olası değerler listesidir veya tüm olası değerleri veya ne sıklıkla meydana geleceğini gösteren fonksiyondur.

En çok bilinen dağılımlardan biride normal dağılımdır. Normal dağılım olası değerlerin reel sayılar hattına uzanması halinde oluşur ve verinin çoğu (%68 civarı) merkezde ortalama civarındadır. Daha ileri hareket ettiğimiz zaman küçük değerler karşımıza çıkar.



Ortalama her zaman normal dağılımda orta kısımdadır ve standart sapma ortalamadan büyüküm noktalarına olan uzaklık ölçülerek bulunur (Büyüküm noktaları ; iç

bükeylikten dış bükeyliğe geçişinolduğu noktalar) Bu grafikte ortalama 0 ve standart sapma 1 dir.

Merkezi Limit Teoremi (Central Limit Theorem)

N tane birbirinden bağımsız rassal değişkenin (random variable) toplamı olan bir rassal değişkenin n sonsuza giderken ortalama değeri etrafında normal dağılıma sahip olacağını söyler. Daha basit bir ifadeyle birbirinden bağımsız pek çok faktörün etki ettiği bir olayın normal dağılıma yakın davranacağını söyleyebiliriz. Bu sayade anakütle dağılımı bilinmeyen değişkenler için, örneklem oluşturarak normal dağılımı kullanabiliriz.

Z- Değerleri (Z-values)

Eğer veri setimiz normal dağılıma sahip ise ve bu tüm veri setini bir standart skora standardize ettiğimizde elde edilen skor veri setimiz için bulduğumuz Z-değeri olur. Başka bir tabirle Z skoru ortalamadan kaç standart sapma uzakta olduğumuzu belirtir. Tüm Z değerleri Standart Normal Dağılım olarak bilinen dağılımdan çıkar ve Standart Normal Dağılım'da ortalama 0'a Standart Sapma is 1'e eşittir.

Yanılma Payı (Margin of Error)

Muhtemelen şu sözü duymuşsunuzdur: "Bu analizde yüzde %05 yanılma payı vardır" Peki nedir bu yanılma payı? Hatasız bir test yapamayacağımız için her testte bir miktar yanılma riskimiz vardır. Bunu 0,05 ; 0,01 ; 0,005 ; 0,0001;... gibi bir düzey olarak benimseyebiliriz. Yanılma payımız küçüldükçe, teste olan güven düzeyimiz yükselir. O nedenle istatistikçiler olabildiğince az yanılma ile test yapmak isterler. Yine de $\alpha = 0,05$ ve $\alpha = 0,01$ düzeyleri en çok kullanılanlardır.

Bu site, hizmet sunmak ve trafiği analiz etmek için Google'a ait çerezleri kullanır. Hizmet kalitesinden emin olunması, kullanım istatistikleri oluşturulması, kötüye kullanımın algılanması ve çözüme kavuşturulması amacıyla performans ve güvenlik metriklerinin yanı sıra IP adresiniz ve kullanıcı aracınız Google ile paylaşılır.

DAHA FAZLA BİLGİ TAMAM

olarak varsayalım. Başlarken popülasyondan bir örneklem aldık (Örneğin Türkiye hanelerinden 1000 tanesi) Ve sonra bu örnekleme ilişkin istatistiklerimizi bulduk. Eğer tüm bir popülasyon hakkında bir yorum yapmak istiyorsak bunun sonrasında bulduğumuz değerlere basit bir tabirle ekleme veya çıkartma yapmamız gerekir çünkü örnekten örneğe değişim görülebilir. Bu ekleme ve çıkartma, örneklem istatistiğinin tahmini amacıyla eklediğimiz parametrenin yanılma payıdır. Ve bunun sonucunda güven aralığımızı elde ederiz. Basit bir Örnekle evden işe 30 dkda gidiyorsunuz ama bu bazen 5 dk erken bazen 5 dk geç oluyor burdaki 5 dk sizin yanılma payınız ve güven aralığınız ise 25 ve 35 arasındadır.

Hipotez Testi (Hypothesis testing)

Hipotez kısaca doğruluğu bir araştırma ya da deney ile test edilmeye çalışılan öngörülere, denencelere denir.

Hipotez testleri bir örneklem ortalaması ile bu örneklemin çekilmiş olduğunu düşündüğümüz ortalaması etrafındaki farkın anlamlı olup olmadığını (yani önemli bir fark olup olmadığını) araştırmamızı sağlayan testlerdir.

Örneğin; Bir pizza şirketi ortalama siparişlerini adreslere 30 dakikada teslim ediyor ve bize bunu test etmemiz isteniyor. Bunun için tüm sipariş sürelerini içeren ana kütleden rastgele örneklem seçimi yaparak ortalama sürenin ortalama periyodun dışına çıkıp çıkmadığını kontrol edip bu iddianın anlamlı olup olmadığını test ederiz. Ayrıca bulduğumuz sonuçlar örneklemden örnekleme farklılık gösterebilceğinden bu hesap değerlerindeki dikkate alınmalıdır.

Örnek bir hipotez;

H_0 : Ortalamalar arasında anlamlı bir farklılık yoktur
 H_1 : Ortalamalar arasında anlamlı bir farklılık vardır

p – Değeri ve Anlamlılık (p-values)

Hipotez testi popülasyon hakkında yapılmış olan iddianın gerçekliğini test etmek için kullanılır. Bu denemede olan iddiaya sıfır hipotezi dersek ve bu hipotezin asılsız sonuca varılması durumunda inanacağımız diğer hipotez ise alternatif hipotez olur. Hipotez testlerinin doğruluğunun gücünü ölçmek için p – değerini kullanırız. P değeri 0 ile 1 arasında olan bir numaradır.

Küçük p – değeri ($p \leq 0,05$) sıfır hipotezini redederiz
Büyük p - değeri ($p > 0,05$) sıfır hipotezini rededemeyiz.

Örneğin biraz önceki örnekten gidelim; bir Pizza şirketi siparişlerini evlere ortalama en fazla 30 dakikada getirdiğini söylüyor, bizde buna inanmıyoruz ve 30 dakikadan fazla olduğunu iddia ediyoruz. Buna göre hipotezimiz;

H_0 : Eve servis süresi max ortalama 30 dk.
 H_a : Eve servis süresi 30 dk'dan daha fazladır.

Bu site, hizmet sunmak ve trafiği analiz etmek için Google'a ait çerezleri kullanır. Hizmet kalitesinden emin olunması, kullanım istatistikleri oluşturulması, kötüye kullanımın algılanması ve çözüme kavuşturulması amacıyla performans ve güvenlik metriklerinin yanı sıra IP adresiniz ve kullanıcı aracınız Google ile paylaşılır.

DAHA FAZLA BİLGİ TAMAM

Kullanılmış Olur.

Gönderen **istatistik** zaman: 07:54

Hiç yorum yok:

Yorum Gönder

Yorum bırakmak için aşağıdaki düğmeyi tıklayarak Google ile oturum açın.

GOOGLE İLE OTURUM AÇ

Sonraki Kayıt

Ana Sayfa

Önceki Kayıt

Kaydol: **Kayıt Yorumları (Atom)**

Hakkımda

istatistik

Profilimin tamamını görüntüle

Basit teması. Tema resimleri **luoman** tarafından tasarlanmıştır. **Blogger** tarafından desteklenmektedir.