

İSTATİSTİK

Fundamentals of Statistics-1

Introduction (Genel Başlangıç)

Öncelikle şu video linki verilmiştir: <https://youtu.be/sxQaBpKfDRk>

Her gün dünyayı tanımlayan veya analiz eden sayısal bilgilerle karşılaşırız. Örneğin, Çin'de korona virüsün ortaya çıkmasından sonra yayınlanan bazı haberler.

"44.000'den fazla doğrulanmış vaka arasında, 9 yaşında ve daha küçük çocuklarda tespit edilen enfeksiyon sayısı çok azdı. - yalnızca 416 vaka başka bir deyişle vakaların yaklaşık %1'i. Ve hiçbir ölmedi.

9 ile 19 yaşları arası çocuklar ve gençler arasındaki vakalar bile nadirdi; bu grupta, tüm çalışma grubunun %1,2'sini temsil eden 549 vaka vardı. O yaş grubunda tek bir ölüm oldu. Buna karşılık, hastalığa yakalanan 80 yaş ve üstü yaklaşık 1.400 kişinin %20'si öldü."

Yeni teknolojiler, insanların giderek artan miktarda veriyi işlemesini ve yorumlamasını sağladığından, istatistikler günümüzde birçok karar alma sürecinin giderek daha önemli bir parçası haline geliyor.

Basit bir ifadeyle istatistik, bilimin grameridir. İşte diğer istatistik tanımları:

Wikipedia'ya göre:

- İstatistik, verilerin toplanması, düzenlenmesi, analizi, yorumlanması ve sunulmasıyla ilgili disiplindir.

Merriam-Webster's Collegiate Dictionary'ye göre:

- İstatistik, sayısal veri yığınlarının toplanması, analizi, yorumlanması ve sunulmasıyla ilgilenen bir matematik dalıdır.

Statistics helps us with decision-making processes.

Select one:

☒ True ✓

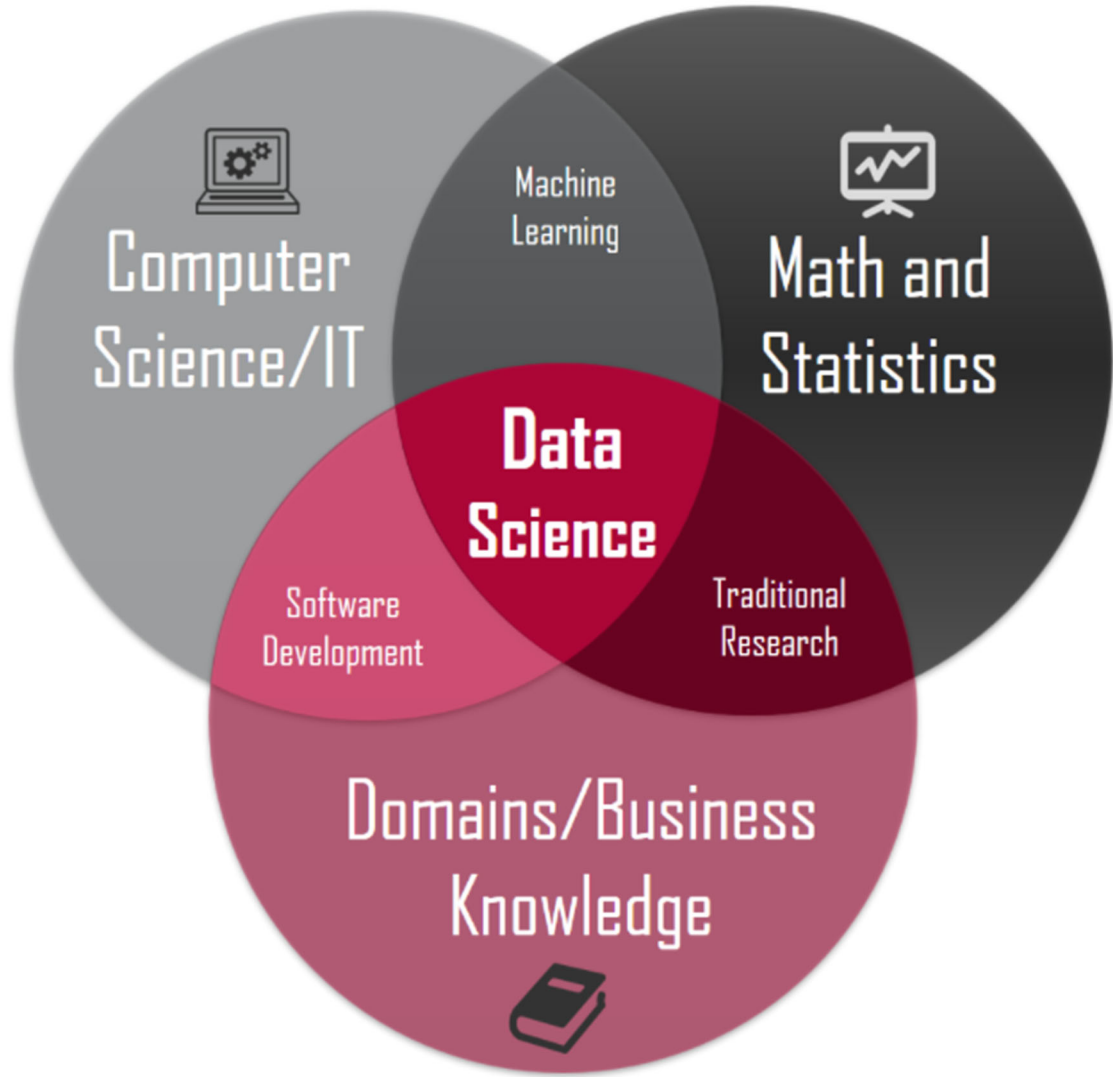
☐ False

Check

Why Should You Learn Statistics (Neden İstatistik Öğrenmelisin?)

Amerikan İstatistik Derneği, istatistikleri "verilerden öğrenme ve belirsizliği ölçme, kontrol etme ve iletme bilimi" olarak tanımlar. Veriler ve istatistikler arasındaki ilişki çok güçlüdür çünkü gerçek dünyadaki veriler istatistiklerle işlenebilir ve anlamlı eğilimler istatistik çerçevesinde matematiksel hesaplamalar yapılarak ondan elde edilebilir.

"Bir programcıdan daha fazla istatistik ve bir istatistikçiden daha fazla programlama bilen bir kişi" tanımı, bir veri bilimcinin meşhur tanımlarından biridir. Öte yandan, veri bilimi bilgisayar bilimi, matematik ve istatistiksel beceriler ve alan bilgisinin kesişme noktasındadır. Aşağıdaki resim, istatistik ve veri bilimi arasındaki ilişkiyi göstermektedir.



Bu nedenle istatistik, veri bilimi dünyasında önemli bir yere sahiptir. Bu derste istatistiksel konuları bir istatistikçi gibi derinlemesine tartışmayacağız. Ancak, istatistiğin temellerini bir bilim adamının bilmesi gerektiği kadar ele alacağız.

The relationship between data and statistics is very strong. Because it can be described as "the science of learning from data."

Select one:

☒ True ✓

☐ False

Check

Types of Data (Data Tipleri)

Introduction to Data (Verilere Giriş)

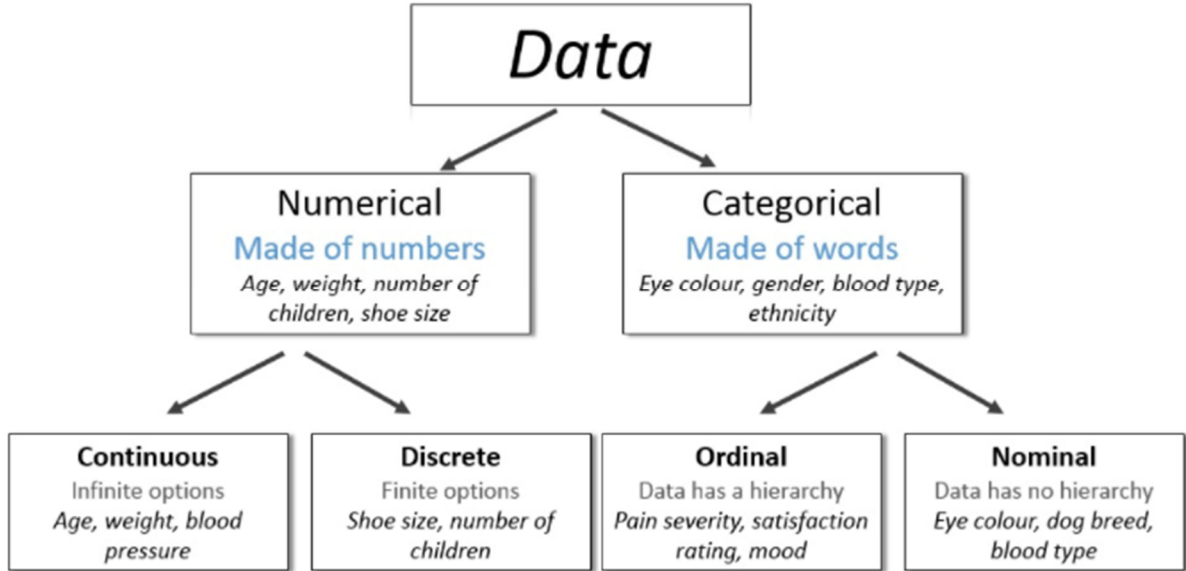
Bilim adamları, titiz yöntemler ve dikkatli gözlemler kullanarak soruları yanıtlamaya çalışırlar. Saha notları, anketler ve deneylerden toplanan bu gözlemler, istatistiksel bir araştırmanın omurgasını

oluşturur ve veri olarak adlandırılır. İstatistik, verilerin en iyi nasıl toplanacağı, analiz edileceği ve bunlardan sonuç çıkarılacağı üzerine yapılan çalışmadır.

Data (Veri)

Veriler, gözlem yoluyla toplanan, genellikle sayısal olan özellikler veya bilgilerdir

Farklı veri türlerini anlamak, Keşif Verileri Analizi (Exploratory Data Analysis) (EDA) yapmak için çok önemli bir ön koşuldur çünkü belirli veri türleri için yalnızca belirli istatistiksel ölçümleri kullanabilirsiniz.



İki ana veri türü vardır: sayısal (numerical) (veya nicel veriler) ve kategorik (categorical) (veya nitel) veriler.

Numerical Data

Sayısal veriler, sayımları veya ölçümleri içerir (counts or measurements). Ayırık (discrete) ve sürekli (continuous) veriler olarak daha da sınıflandırılabilir.

Ayrık veriler yalnızca sonlu değerler içerir. Örneğin bir okuldaki öğrenci sayısı, bir şehirdeki bina sayısı, bir sınıftaki sıra sayısı vb. Bir okuldaki olası öğrenci sayısı sonsuz olamaz. 300, 301 veya 302 olabilir, ancak 300 ile 301 arasında bir şey olamaz. Alabileceği değerleri bir şekilde sayabilirsiniz.

Sürekli veriler, herhangi iki değer arasında sonsuz sayıda değere sahip olabilir. Örneğin bir kişinin yaşı, boyu veya kilosu, özel bir göreve atanan zaman, ev fiyatı, hisse senedi fiyatı vb. Bir hisse senedi fiyatı 10, 11 veya 10.4565 dolar olabilir, bu nedenle sonsuz sayıda değere sahip olabilir. Bir kişinin yaşı sürekli bir veridir çünkü tam yaştan 28, 28.1 veya 28.5 olabilir.

Ayrık ve Sürekli Veriler

Olası değerleri 0, 1, 2, 3, ... gibi bir dizi ayrı sayı oluşturuyorsa sayısal bir değişken ayrıktır.

Olası değerleri bir aralık oluşturuyorsa sayısal bir değişken sürekli. (A numerical variable is continuous if its possible values form an interval.)

There are two main types of data types: **numerical** and **categorical** ✓

Check

- categorical
- discrete
- continuous

Categorical Data

Kategorik veriler yalnızca kadın / erkek cinsiyetler, otomatik / yarı otomatik veya manuel vites kutuları gibi bir dizi olası kategoriye temsil eden belirli bir değer kümesini alabilir.

Kategorik verilere sayılar atayabilirsiniz, örneğin erkek için 0, kadın için 1 kullanabilirsiniz. Ancak bu durumda bile sayıların gerçek sayısal anlamı yoktur. Kategorileri temsil ederler.

Kategorik veriler ayrıca nominal (nominal) ve sıralı (ordinal) veriler olarak sınıflandırılabilir. Nominal veriler, bir sipariş verilmeden basitçe bir şeyi adlandırır. Dişi / erkek cinsiyetler, otomatik / yarı otomatik veya manuel vites kutuları nominal verilere örnektir.

Nominal verilerden farklı olarak, sıra verileri bir sipariş gerektirir. Örneğin, bir restoranda akşam yemeği yediğinizi ve ankete katılıp katılamayacağınızı sorduklarını varsayalım. Yemeğin tadını kötü için "1", ortalama "2" ve iyiye "3" olarak sıralayabilirsiniz. Anket verileri, sıralı verilere iyi bir örnektir. Burada kategorik veriler için sayılar kullanıyoruz, ancak yukarıda bahsettiğimiz gibi, bunların gerçek sayısal anlamı yoktur. Bir gömlek için vücut ölçüleri ile ilgili başka bir örnek verilebilir. Küçük, Orta, Büyük boyutların tümü sıra verilerini temsil eder. Çünkü bu kategoriler arasında bir düzen var.

Ordinal/Sıralı ve Nominal Veriler

Sıralı veri, kategorilerin bir anlamda düşükten yükseğe doğru sıralandığı kategorik bir veridir.

Nominal veri, doğal bir düzeni veya sıralaması olmayan kategorik bir veridir.

Which of the following statements are true?

Select one or more:

- ☒ I. All variables can be classified as quantitative or categorical variables. ✓ True!
- ☐ II. Categorical variables can be continuous variables.
- ☒ III. Quantitative variables can be discrete variables. ✓ True!

Check

Check Yourself Soruları:

Which of the following is NOT a continuous data?

Select one:

- ☐ The volume of water in a swimming pool each day
- ☐ A person's height each year
- ☐ A person's weight on each birthday
- ☒ Cars finished in a factory each day ✓ Congrats! You are right.

Which of the following is NOT a discrete data?

Select one:

- ☐ How many students are absent from school each day
- ☐ How many widgets a business sell each day
- ☐ The number of people who drive through a red light each hour during rush hour
- ☒ Height of a sunflower as measured each day ✓ Congrats! You are right.

"Hours Studying" is an example of categorical data.

Select one:

- ☐ True
☒ False ✓

"Agree / Disagree" is an example of categorical data.

Select one:

- ☒ True ✓
☐ False

Level of Measurements (Ölçüm Seviyesi)

Ölçme düzeyi veya ölçü skalası, değişkenlere atanan değerler içindeki bilginin özünü açıklayan bir sınıflandırmadır.

Değişken (Variable)

Değişken, bir çalışmada gözlemlenen herhangi bir özelliktir.

Dört farklı ölçüm seviyesi vardır: nominal (nominal), sıralı (ordinal), aralık (interval) veya oran (ratio). Önceki dersten hatırlayacağımız gibi, nominal ve sıralı veriler kategorik verilerin iki alt kategorisidir. Buna uygun olarak, kategorik veriler için nominal veya sıralı ölçüm seviyelerinden birini ve sayısal veriler için ölçümlerin aralık veya oran seviyesini kullanacağız. Burada önemli olan nokta, bu seviyeler arasında bir hiyerarşi olması ve bu hiyerarşinin nominalden oran seviyesine yükselmesidir.

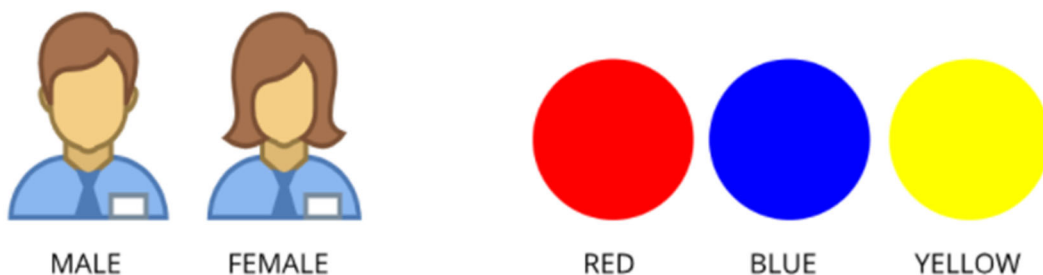
Bir veri bilimcisinin farklı ölçüm düzeylerini anlaması gerekir çünkü bu, verilerimiz için hangi istatistiksel analizin uygun olduğunu belirler.

We use one of the nominal or ordinal levels of measurements for categorical data, and interval or ratio level of measurements for numerical data.

Check

Nominal Level of Measurement (Nominal Ölçüm Seviyesi)

İlk seviye, nominal ölçüm seviyesidir. Daha önce de belirttiğimiz gibi, nominal düzey, bir sıra atamadan basitçe bir şeyi adlandırır. Bu düzeyde, değişkendeki sayılar yalnızca verileri sınıflandırmak için kullanılır. Örneğin erkekler için 0 ve kadın için 1 kullanmak Erkek/Kadın, Kırmızı/Mavi/Sarı veya Başarılı/Başarısız sınıflandırmaları, nominal ölçüm düzeyi için örneklerdir.



Tips:

Nominal Level of Measurement simply names something without assigning an order.

Automatic and manual car classification is **not** an example of the nominal level of measurement.

Select one:

☐ True

☒ False ✓

Ordinal Level of Measurement (Sıralı Ölçüm Düzeyi)

İkinci seviye, sıralı ölçüm seviyesidir. Bu seviyede, nitelikler sıralanır, ancak nitelikler arasındaki mesafelerin (distance) bir anlamı yoktur. Restorandaki anketi hatırlayın. Bu ölçüde, daha yüksek sayılar daha fazla lezzet anlamına gelir. Ancak 1'den 2'ye kadar olan mesafe 2'den 3'e kadar aynı olmayabilir. Sıralı ölçüm seviyesinde, değerler arasındaki aralık yorumlanamaz.



HOT



HOTTER



HOTTEST

Tips:

In the ordinal level, the attributes are ordered.

But distances between attributes have no meaning.

Small-Medium-Large shirt classification is an example of the ordinal level of measurement.

Select one:

☒ True ✓

☐ False

Interval Level of Measurement (Aralık Ölçüm Seviyesi)

Sıralı ölçekten farklı olarak, ölçüm aralığı düzeyinde, nitelikler arasındaki mesafenin anlamı vardır. Örneğin sıcaklık ölçülürken 40-50 derece arasındaki mesafe, 60-70 derece arasındaki mesafe ile aynıdır. Değerler arasındaki aralık yorumlanabilir. Bu, sıralı ölçekler için çalışmadığı bir aralığın ortalamasını hesaplamamanın mantıklı olduğu anlamına gelir. Bununla birlikte, ölçüm aralığı seviyesinde oran (ratio) herhangi bir anlam ifade etmiyor. 100 derecenin 50 derecenin iki katı sıcak olduğunu söylemek mümkün değil.

Tips:

The distance between attributes does have meaning.

But the ratio does not make any sense.

Temperature (Fahrenheit), Temperature (Celsius), SAT score (200-800) and credit score (300-850) are good examples for interval scale.

Select one:

☒ True ✓

☐ False

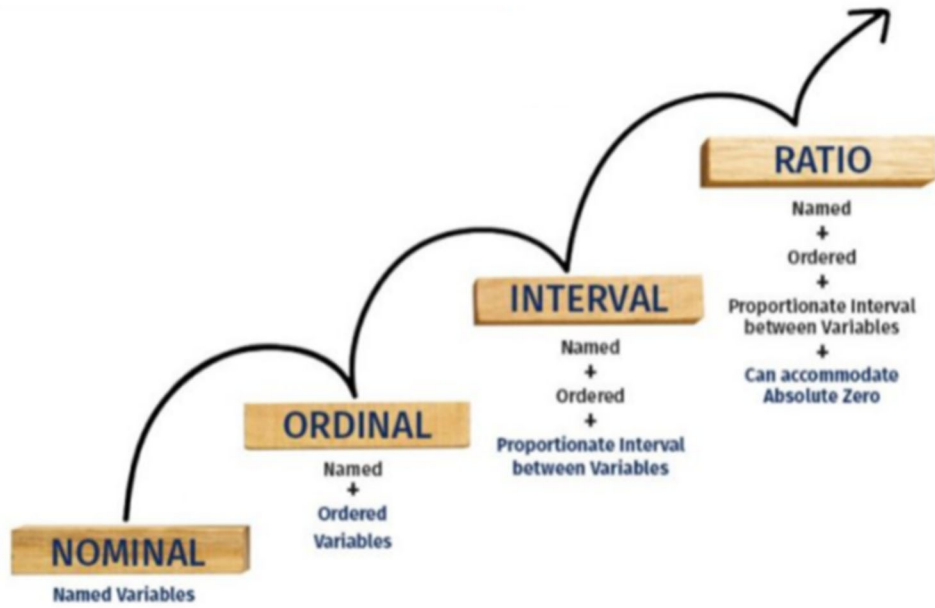
Ratio Level of Measurement (Oran Ölçüm Seviyesi)

Dördüncü seviye, ölçüm oranı seviyesidir. Bu düzeyde, gözlemler aynı aralıklara sahip olmanın yanı sıra sıfır değerine de sahiptir. Sıfır değeri, bu ölçüm düzeyini aralık ölçeğinden farklı kılar. Bu ölçüm seviyesine örnekler arasında uzunluk, boy, ağırlık veya cep telefonu şarj kapasitesi bulunur. Aralıklı ölçeklerden farklı olarak oranlar anlamlıdır. Bir nesnenin iki kat yüksekliğe sahip olduğunu söylemek mümkündür. Çünkü sıfır noktasına sahip olmak onu anlamlı kılar.

Tips:

Ratios make sense.

There is a zero point.



Tamamlayıcı video linki <https://youtu.be/hZxnzft5v8>

What is your weight in kilograms?

The question above is an example of ratio level measurement.

Select one:

☒ True ✓

☐ False

Graphical Representation of Data (Verilerin Grafik Gösterimi)

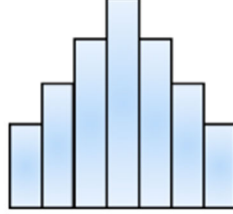
Data Patterns in Statistics (İstatistiklerdeki Veri Modelleri)

Grafik verilerdeki kalıpları görüntülemek için kullanışlılık sunar. Genel olarak verilerdeki kalıpları şu şekilde tanımlayabiliriz: merkez (center), yayılmış (spread), şekil (shape) ve olağandışı (unusual) özellikler.

Bilinen bazı dağıtımların simetrik (symmetric), çan şeklinde (bell-shaped), çarpık (skewed) vb. gibi bazı özel tanımlayıcı etiketleri vardır.

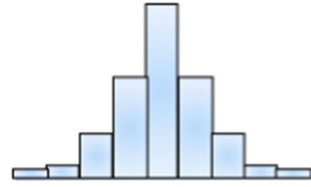
Center (Merkez)

Bir dağılımın merkezi, grafiksel olarak dağılımın medyanında bulunur. Böyle bir grafik, gözlemlerin neredeyse yarısının her iki tarafta olduğunu gösterir. Her bir sütunun yüksekliği, gözlemlerin sıklığını gösterir.

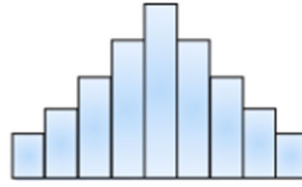


Spread (Yayılmış)

Bir dağılımın yayılmasından bahsederek, verilerin varyasyonu anlaşılır. Gözlem grubu geniş bir yelpazeye yayılırsa, yayılma daha büyüktür. Gözlemler daha dar bir aralıkta tek bir değer etrafında ortalanırsa, dağılım daha küçüktür.



Less Spread

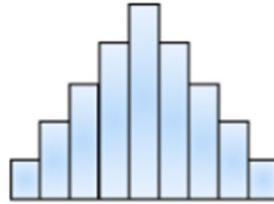


Wide Spread

Shape (Şekil)

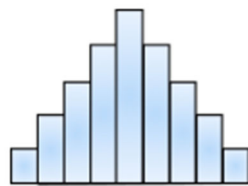
Bir dağılımın şekli aşağıdaki özellikler kullanılarak tanımlanabilir.

Symmetry (Simetri): Simetrik dağılımda, grafik, her iki yarısı diğerinin ayna görüntüsü olacak şekilde merkezde bölünebilir. Normal Dağılımda dağılımın iki tarafı eşit ve simetrik.

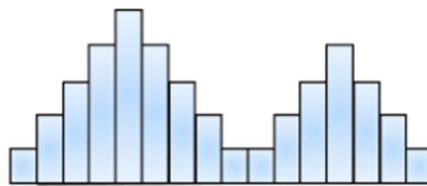


Symmetric

Number of peaks (Zirve sayısı): Bir veya birden fazla tepe noktası olan dağılımlar. Bir net tepe ile dağıtım tek modlu olarak bilinir ve iki açık tepe ile dağıtım iki modlu olarak bilinir. Merkezde tek bir tepe simetrik dağılım, çan şeklinde olarak adlandırılır.

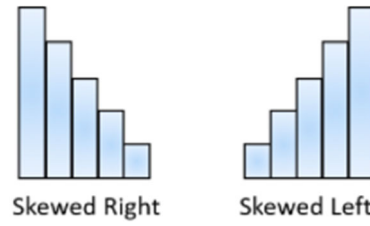


Unimodal,
Bell Shaped

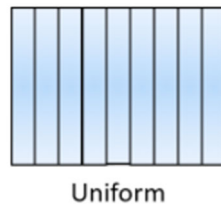


Bimodal

(Skewness) Çarpıklık: Bazı dağılımlar, grafiğin bir tarafında diğer tarafında olduğundan daha fazla gözleme sahip olabilir. Daha yüksek değerlere yönelik daha az gözlemi olan dağılımların sağa çarpık olduğu söylenir; ve daha düşük değerlere doğru daha az gözlem içeren dağılımların sola çarpık olduğu söylenir.



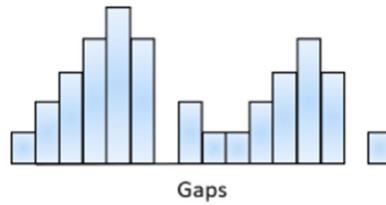
(Uniform) Tekdüze: Gözlem kümesinin zirvesi olmadığında ve dağılım aralığı boyunca eşit olarak yayılmış veriye sahip olduğunda, dağılıma tekdüze dağılım denir.



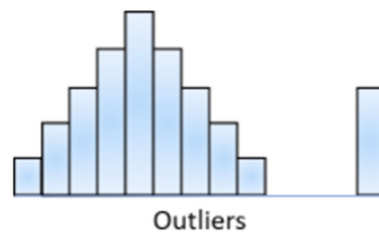
Unusual Features (Olağandışı Özellikler)

Veri modellerinin yaygın alışılmadık özellikleri, boşluklar (gaps) ve aykırı (outliers) değerlerdir.

(Gaps) Boşluklar: Boşluklar, bir dağılımın gözlemi olmayan alanlarına işaret eder. Dağılımın ortasında herhangi bir gözlem olmadığı için aşağıdaki figürde boşluk var.



(Outliers) Aykırı Değerler: Dağılımlar, diğer gözlem verileri kümesinden büyük ölçüde farklı olan uç değerlerle karakterize edilebilir. Bu uç değerler, aykırı değerler olarak adlandırılır. Aşağıdaki şekil, aykırı değere sahip bir dağılımı göstermektedir.



Distributions may be characterized by extreme values that differ greatly from the other set of observation data. These extreme values are referred as .

Check

skewness

outliers

gaps

Frequency Table (Frekans tablosu)

Sıklık tablosu, ögeleri listeleyen ve ögelerin kaç kez meydana geldiğini gösteren bir tablodur. Bir kategorinin veya sayısal bir değerin göreceli frekansı, toplam veri sayısına bölünen ilişkili frekanstır. Aşağıdaki tablo, çeşitli yanıt kategorilerinin frekanslarını göstermektedir. Ayrıca, her kategorideki yanıtların oranı olan göreceli sıklıkları da gösterir. Örneğin, "Ön Uç Geliştirici" için göreceli frekans $25/100 = 0,25$ 'tir.

Developer Type	Frequency	Relative Frequency
Front-end Developer	25	0.25
Backend Developer	15	0.15
Full-stack Developer	20	0.20
Data Scientist	40	0.40

A frequency table is a listing of possible values for a variable, together with the number of observations for each value.

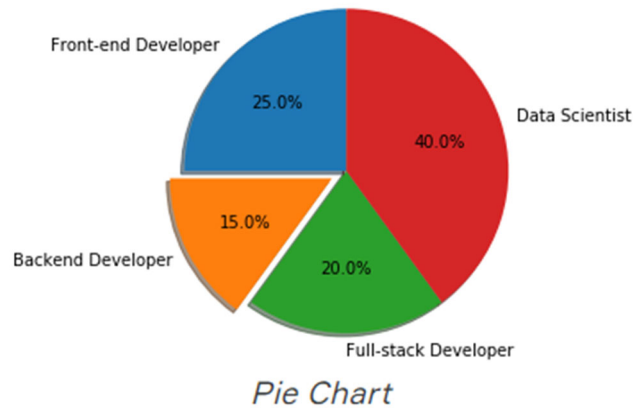
The **relative frequency** of a category or a numerical value is the associated frequency divided by the total number of data.

Pie Chart (Pasta diyagramı)

Pasta grafiğinde, her kategori pastanın bir dilimi ile temsil edilir. Dilimin alanı, kategorideki yanıtların yüzdesi ile orantılıdır. Bu, 100 ile çarpılan göreceli frekans.

A pie chart is a circle having a "slice of the pie" for each category.

Aşağıdaki pasta grafikten de görülebileceği gibi, şirketteki geliştiricilerin çoğu Data Scientists (%40), geliştiricilerin %25'i Front-end Developers, %20 Tam Full-stack Developers ve %15 Backend Developers.



Pasta grafikler, az sayıda kategorinin göreceli frekanslarını görüntülemek için etkilidir. Bununla birlikte, çok sayıda kategorinin olduğunda önerilmezler. Pasta grafikler, iki farklı anket veya deneyin sonuçlarını karşılaştırmak için kullanıldıklarında da kafa karıştırıcı olabilir.

Tips:

When slices become too small, pie charts have to rely on colors, textures or arrows so the reader can understand them.

Avoid!

Pie charts are not recommended if there are too many categories of data being presented.

When there are many categories, it is best to use a pie chart.

Select one:

☐ True

☒ False ✓

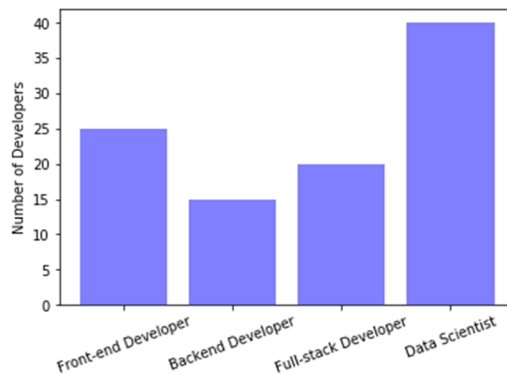
Bar Chart (Grafik çubuğu)

Çubuk grafikler, farklı kategorilerin frekanslarını temsil etmek için kullanılabilir. Çubuk grafik, bir grafiğe çizilen sütunlardan oluşur.

Bar Chart (Grafik çubuğu)

Sütun grafik, temsil ettikleri değerlerle orantılı yüksekliklere sahip dikdörtgen çubuklarla kategorik verileri temsil eden bir grafik türüdür.

Bir şirkette çalışan geliştiricilerin bir çubuk grafiği aşağıda gösterilmiştir. Frekanslar Y ekseninde gösterilir ve geliştiricilerin türü X ekseninde gösterilir. Tipik olarak, Y eksen, pasta grafiklerde olduğu gibi gözlemlerin yüzdesinden ziyade her kategorideki gözlemlerin sayısını gösterir.



Çubuk grafiğin nasıl okunacağı aşağıda açıklanmıştır.

- Sütunlar, kategorik bir değişkeni temsil eden bir etiketin üzerine yerleştirilmiştir.
- Sütunun yüksekliği, sütun etiketi ile tanımlanan grubun boyutunu belirtir.

İpuçları:

Çubuklar dikey veya yatay olarak çizilebilir.

Dikey çubuk grafik bazen sütun grafik (column chart) olarak adlandırılır.

Avoid!

Do not start the axis with a value above zero. If you do this, you shrink the bars and get confusing visuals.

Do not use the rainbow effect. Although coloring the graphics sounds good, it often makes the tables difficult to understand.

Three-dimensional bar charts make it easier to interpret the data.

Select one:

☐ True

☒ False ✓

Histogram

Histogram, bir dağılımın grafik temsidir.

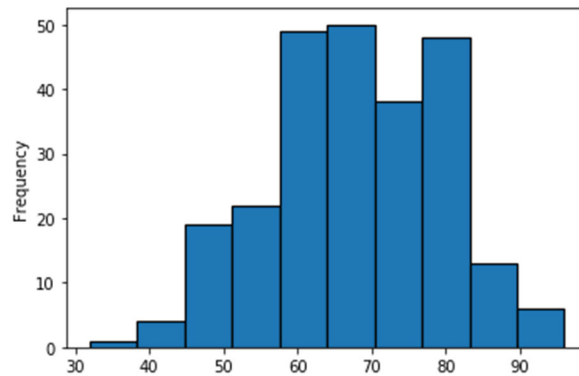
Histogram, bir dağılımın şeklini görüntülemek için grafiksel bir yöntemdir. Çok sayıda gözlem olduğunda özellikle yararlıdır. İstatistik testinde 250 öğrencinin puanlarından oluşan bir örnek başlıyoruz. Öğrencilerin puanları 32 ile 96 arasında değişiyordu.

İlk adım, bir sıklık tablosu (frequency table) oluşturmaktır. Ne yazık ki, 100'den fazla satır içeren basit bir frekans tablosu çok büyük olacaktır. Tabloyu basitleştirmek için, puanları aşağıdaki tabloda gösterildiği gibi gruplandırıyoruz.

Interval's Lower Limit	Interval's Upper Limit	Class Frequency
32	38.4	1
38.4	44.8	4
44.8	51.2	19
51.2	57.6	22
57.6	64	49
64	70.4	50
70.4	76.8	38
76.8	83.2	48
83.2	89.6	13
89.6	96	6

Bu tabloyu oluşturmak için puan aralığı, sınıf aralıkları (intervals) adı verilen aralıklara bölündü. İlk aralık 32'den 38.4'e, ikincisi 38.4'ten 44.8'e, vb. Daha sonra, sınıf frekanslarını elde etmek için her aralığa düşen puanların sayısı sayıldı. İlk aralıkta bir puan, ikincide 4 puan vb. vardır. Sınıf aralıklarının genişlikleri (widths), bazen (bin widths) "bin" genişlikleri olarak adlandırılır. "Bin" genişliği seçiminiz, sınıf aralıklarının (class intervals) sayısını belirler.

Bir histogramda, sınıf frekansları çubuklarla temsil edilir. Her çubuğun yüksekliği, sınıf frekansına karşılık gelir. Bu verilerin histogramı aşağıda gösterilmektedir.

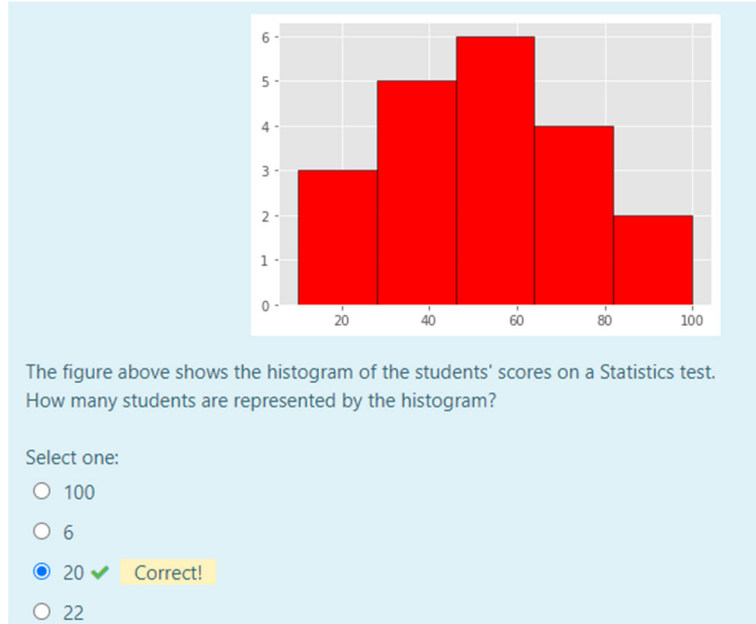


Histogram of scores on a Statistics test.

Histogram, puanların çoğunun, aşırı uçlarda daha az puanla, dağılımın ortasında olduğunu açıkça ortaya koymaktadır. Grafikten, frekansın 64 ila 70,4 puan grubunda en yüksek olduğunu görebilirsiniz.

Tips:

With bar charts, the labels on the X axis are categorical; with histograms, the labels are quantitative.



Population & Sample (Nüfus ve Örnek)

Population (Nüfus)

Nüfus, incelenen tüm öğelerin bir veya daha fazla ortak özelliklerle bir araya getirilmesi anlamına gelir. Nüfus sadece insanlarla sınırlı değildir, ayrıca hayvanlar, nesneler, binalar, arabalar vb. Nüfusu oluşturabilir.

Population

A population is the total group about whom you want to make conclusions.

Örneğin ABD'de yaşayan tüm insanlar, bir ülkedeki tüm binalar, Hindistan'daki engelli çocuklar, dünyadaki tüm öğrenciler, Avrupa'da dizel arabalar, Amerika'daki tüm koyunlar bir nüfus oluşturabilir. Herhangi bir boyutta olabilir ve bir popülasyondaki üye veya üye sayısı, popülasyon büyüklüğü olarak bilinir. Bir ülkede yüz milyon insan varsa, o zaman nüfus büyüklüğü 100 milyondur.

Sample (Örneklem)

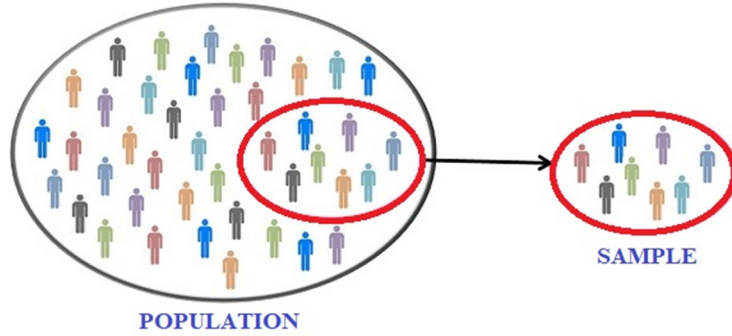
İstatistiksel bir çalışma yürütülürken, örnekler esas olarak popülasyon büyüklüğü popülasyonun tüm üyelerini içermeyecek kadar büyük olduğunda kullanılır. Örneğin 100 milyon kişiden oluşan bir araştırma yapmaktansa, 100 milyon kişiden seçilen 2000 kişi ile araştırma yapmak daha kolaydır. Popülasyondan seçilen katılımcılar bir örnek teşkil eder ve seçim süreci örnekleme olarak bilinir. İncelenen öğelere örnekleme birimleri denir ve bir örnekteki birim sayısına örnek boyutu denir.

Sample

A sample is a subset of the population for whom you actually have data.

Özetle; nüfus, belirli özelliklerle tüm üyeleri temsil eder. Öte yandan, örnek popülasyonun özel bir işlemle seçilen sınırlı bir alt kümesidir. Nüfus her zaman araştırmanın hedefidir. Koleksiyondan örnek alarak popülasyon hakkında bilgi ediniriz.

Aşağıdaki resim popülasyon ve örneklem arasındaki ilişkiyi çok iyi açıklamaktadır.



Population vs. Sample

İpuçları:

Örnekleri gözlemliyoruz ama popülasyonlarla ilgileniyoruz.

Sample is a limited subset of the Population , selected by a special process.

Check

Population Parameters and Sample Statistics (Popülasyon Parametreleri ve Örnek İstatistikler)

Şimdiye kadar sunulan sayısal özetler arasında, ortalama \bar{x} ve standart sapma s , pratikte en yaygın kullanılanlardır. Bunları metnin geri kalanında sık sık kullanacağız. \bar{x} ve s 'yi tanımlayan formüller örnek verilere atıfta bulunur. Örnek istatistiklerdir.

Örnek istatistikleri ve popülasyon için karşılık gelen parametre değerleri arasında ayırım yapacağız. Popülasyon ortalaması, popülasyondaki tüm gözlemlerin ortalamasıdır. Popülasyon standart sapması, popülasyon ortalaması hakkındaki popülasyon gözlemlerinin değişkenliğini tanımlar. Bunlar genellikle bilinmez. Çıkarımsal istatistiksel yöntemler, örnek istatistiklere dayanarak nüfus parametreleri hakkında kararlar ve tahminler yapmamıza yardımcı olur.

Tips:

A parameter is a numerical summary of the population, and a statistic is a numerical summary of a sample.

Population attributes are called parameters ☒

Sample attributes are called statistics ☒

Check Yourself Soruları 2:

Decide the Level of Measurement.
"The teacher of a class records the eye color of each student."

Select one:

☒ nominal ☒ Congrats! You are right.

☐ interval

☐ ordinal

☐ ratio

Decide the Level of Measurement.

"A classroom teacher records the math grade (A, B, C, D, E) of each student."

Select one:

- ☒ ordinal ✓ Congrats! You are right.
- ☐ interval
- ☐ ratio
- ☐ nominal

Decide the Level of Measurement.

"A meteorologist compiles a list of temperatures in degrees Celsius for the month of April."

Select one:

- ☐ nominal
- ☐ ordinal
- ☒ interval ✓ Congrats! You are right.
- ☐ ratio

Decide the Level of Measurement.

"A classroom teacher records the weight of students in her class."

Select one:

- ☐ interval
- ☐ nominal
- ☒ ratio ✓ Congrats! You are right.
- ☐ ordinal

Ministry of Education randomly selected 270 teachers to find out which technology resource is the most effective. 50 teachers chose Safari Montage, 40 selected Learn Zillion, 120 chose Ed Puzzle, and 60 chose Kahoot. Ministry of Education concluded that all teachers prefer Ed Puzzle.

Identify the Population.

Select one:

- ☐ 60 teachers
- ☐ 270 teachers
- ☒ All teachers ✓ Congrats! You are right.
- ☐ 120 teachers

A restaurant wants to know whether customers buy dessert when they eat out. As people leave the restaurant one evening, 15 people are surveyed randomly. 10 people say they usually order dessert when they eat out. The restaurant concluded that most customers order dessert.

Identify the sample.

Select one:

- ☐ dessert
- ☐ all customers
- ☒ 15 customers ✓ Congrats! You are right.
- ☐ 10 customers