**İSTATİSTİĞE GİRİŞ**

# İÇİNDEKİLER:

* **Bazı Tanım ve Kavramlar………………………………………………………………………………… 3 – 9**
* **Sıklık Çizelgeleri ……………………………………………………………………………………………. 10-11**
* **Grafikler ………………………………………………………………………………………………………… 12-16**

A picture containing timeline

Description automatically generated

İstatistik çok sayıda sebep ve değişkene bağlı olayları inceleyen bilim dalıdır. İstatistiğin konusu **yığın** olaylardır. Örneğin; işsizlik bir yığın olaydır ya da boşanma bir yığın olaydır.

Tanımlayıcı İstatistiğe örnek: Seçimlerde partilerin aldığı oyların yüzdelik olarak gösterilmesi ya da bir okuldaki öğrencilerin cinsiyet dağılımının grafikle gösterilmesi vb.

Çıkarımsal (Çıkarsamalı) İstatistiğe örnek: Seçim öncesi yapılan anketlerle seçim sonuçlarının tahmini ya da bir fabrikada üretilen ürünlerin kusurlu olma olasılığını örnek yardımıyla tahmin edilmesi vb.

# BAZI TANIM VE KAVRAMLAR

**ARAŞTIRMA:** Herhangi bir konuda sorunların incelenmesi için yapılan çalışmalardır. Araştırmanın aşamaları:

1. Araştırma konusunun saptanması
2. Araştırmanın planlanması
3. Araştırmanın uygulanması
4. Araştırmanın sunulması

**KİTLE:** Araştırma kapsamına giren aynı özelliğe sahip birimlerin tümüne denir. Örneğin; nüfus sayımında ülkede yaşayan herkes kitledir ya da Türkiye’deki göller bir kitledir.

A picture containing diagram

Description automatically generated

Sonlu Kitleye Örnek: Ankara Üniversitesi İstatistik Lisans Öğrencileri

Sonsuz Kitleye Örnek: Uzaydaki yıldızlar

A picture containing diagram

Description automatically generated

Örnek: Türkiye’deki üniversite öğrencilerinin internet kullanma alışkanlıkları

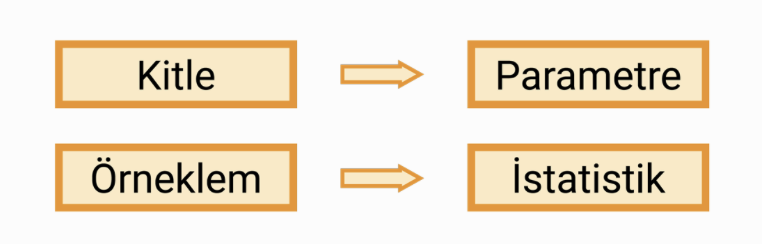
Hedef Kitle: Tüm üniversite öğrencileri

Ulaşılabilir Kitle: 7 bölgeden seçilen birer üniversitedeki (toplam 7 üniversite) kayıtlı öğrenciler.

**ÖRNEKLEM (SAMPLE):** Bir kitleden örnekleme yöntemleri ile seçilen aynı özelliğe sahip birimlerin oluşturduğu topluluktur. Örneklem kitlenin alt kümesidir ve kitleyi temsil edebilmelidir. Bir örneklem maliyeti artırmayacak kadar küçük, istatistiksel olarak anlamlı olacak kadar büyük olmalıdır.

**GÖZLEM (DENEK):** Kitlede ve örneklemde yer alan her bireye/birime denir. Kitle de denek sayısı “N”, örneklemde denek sayısı “n” ile gösterilir.

**PARAMETRE VE İSTATİSTİK:** Kitlenin sayısal ifadelerine **“Parametre”**, örneklemin sayısal ifadelerine **“İstatistik”** denir. İstatistikler kitle özelliklerini (parametrelerini) tahmin için kullanılıyorsa tahmin adını alır. Örnek: Kitle varyansı, Örneklem ortalaması vb.



**DEĞİŞKEN (VARIABLE):** Nicel veya nitel anlamda birimden birime ya da gözlemden gözleme değişen farklılıklara verilen addır. Örneğin; boy uzunluğu, yaş, cinsiyet vb.

A screenshot of a cell phone

Description automatically generated

**A picture containing diagram

Description automatically generated**

**A picture containing diagram

Description automatically generated**

Bağımsız değişkenler, diğer değişkenlerden etkilenmeyen ama onları etkileyen değişkenlerdir. Bağımsız değişken tarafından etkilenen değişken ise bağımlı değişken olarak adlandırılır. Bağımlı ve bağımsız değişken içeren modellerde temel amaç bağımsız değişkendeki değişimin bağımlı değişken üzerindeki etkisini ölçmektir.

**Table

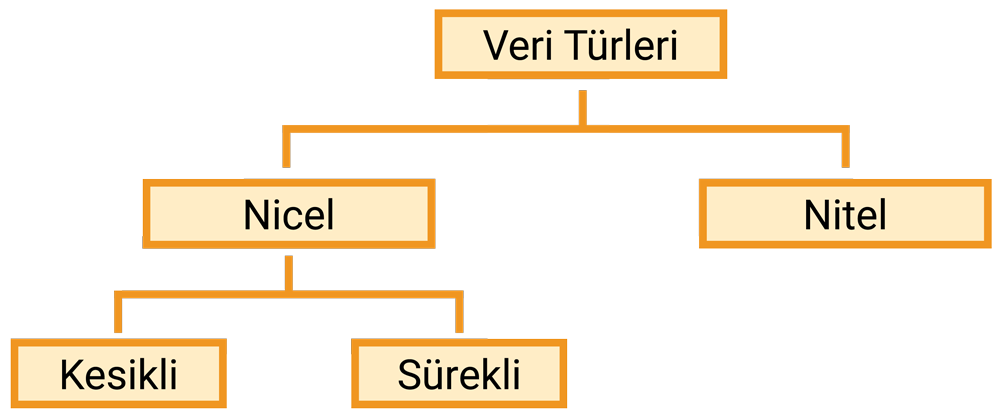
Description automatically generated**

**VERİ (DATA):** Araştırma kapsamında incelenen değişkenlerin ölçüm ya da gözlem değerleridir. Başka deyişle gözlemlerden elde edilen sayısal ya da sayısal olmayan sonuçlardır.

Ham (Raw) Veri: Verilerin düzenlenmemiş haline denir.

Bazı veri toplama türleri:

* Anket (Questionnaire): Anketler örnekleme uygulandığında **araştırma**, kitlenin tamamına uygulandığında **tamsayım (census)** adını alır.
* Deneysel Çalışma (Experimental Study)
* Gözlemsel Çalışma (Observational Study)
* Düzenli olarak tutulan kayıtlar



1. **Nicel Veriler:** Gözlemin sayılabilir özelliklerini veren veri türüdür.

Sürekli Nicel: Ondalık değerler alabilen verilerdir.

Örneğin; Boy uzunluğu, kilo vb.

Kesikli Nicel: Sayılarak elde edilen ve tam sayı değerler alan veri

Örneğin; Çocuk sayısı, Meyve sayısı vb.

1. **Nitel Veriler:** Gözlemin niteliğini belirleyen verilerdir.

Sınıflanabilir Nitel: Ülkeler, cinsiyet vb.

Sıralayabilir Nitel: Başarı durumu, eğitim düzeyi vb.

**ÖLÇME-ÖLÇEK:** Araştırma ile ilgili sayısal değerler elde etmeye ölçme denir. Sonuçta elde edilen bilgiye ölçüm denir. Literatürde ölçek türü yerine veri türü ifadesi de kullanılmaktadır.

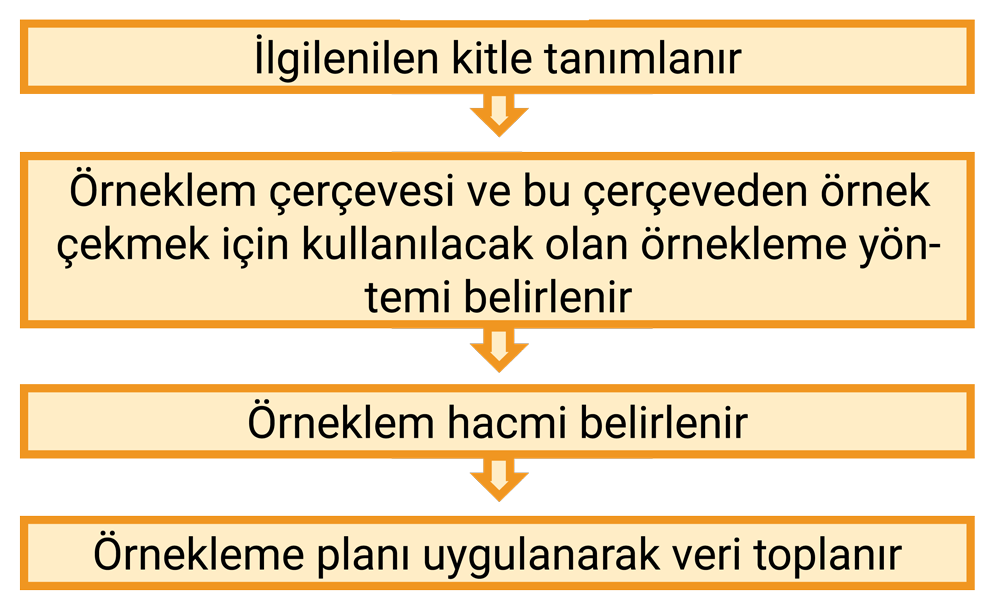
Diagram

Description automatically generated

1. **Sınıflayıcı (Nominal):** Cinsiyete göre sınıflama, hayvanları türüne göre sınıflama gibi.
2. **Sıralayıcı (Ordinal):** Kişilerin eğitim durumu, askeri rütbeler gibi.
3. **Aralıklı (Interval):** Sıcaklık ölçümleri, sınav puanlandırma yöntemleri gibi.
4. **Oransal (Ratio):** Uzaklık ya da ağırlık ölçüsü gibi.

**ÖRNEKLEME (SAMPLING):** Kitleden örneklem seçme işlemine örnekleme denir. Zaman, işgücü ve maliyet avantajı sağlar.

Kitleden örneklem çekmek için belli bir prosedür takip edilir. Bu prosedürün aşamaları aşağıda gösterildiği gibidir.



A screenshot of a cell phone

Description automatically generated

* Olasılığa bağlı olan örnekleme yöntemlerinde kitleyi temsil edebilme yeteneği daha yüksek olan örneklem oluşturulabilir.
* Olasılığa bağlı yöntemler, araştırmacılar tarafından daha çok tercih edilir. Ancak özellikle sosyal bilimler gibi alanlarda olasılığa bağlı olmayan yöntemleri kullanmak hem daha uygun hem de daha yaygındır.
* Olasılığa bağlı yöntemler, olasılığa bağlı olmayan yöntemlere göre daha masraflı ve zaman alıcıdır.

1. **Olasılığa Bağlı Olmayan Örnekleme Yöntemleri:**

Kitledeki birimlerin, örnekleme rastgele olarak çekilmediği ya da bir başka deyişle kitledeki bazı bireylerin örnekleme çekilme olasılıklarının sıfır olduğu yöntemler, olasılığa bağlı olmayan örnekleme yöntemleri olarak isimlendirilir.

Olasılığa bağlı olmayan yöntemlerde kitledeki birimlerin örnekleme çekilme olasılıkları bilinmez, örneklemdeki birimlerin çekimi, keyfi bir şekilde genellikle araştırmacının kişisel değerlendirmeleri ışığında yapılır. Olasılığa bağlı olmayan yöntemlerde rastgele örnekleme hatasını (random sampling error) ölçmek için uygun bir istatistiksel teknik yoktur. Ayrıca örneklemden yola çıkarak kitlenin belli bir karakteristik özelliği hakkında yargıya varmak da oldukça zordur.

Örnek: Bir deterjan firması, üretmiş olduğu X marka deterjanın kullanıcılar tarafından beğenilip beğenilmediğini araştırmak istiyor. Bu amaçla, bir semtteki tüm evler ziyaret edilip kullanıcılara ürün hakkındaki görüşleri soruluyor. İki yetişkinin yaşadığı bir evde genellikle bu soruya çamaşır yıkama işiyle uğraşan kişi cevap vereceğinden diğer kişinin örnekleme seçilme olasılığı sıfırdır.

1. Amaca Bağlı Örnekleme: Araştırıcının kendi görüş ve tecrübelerine dayanarak seçmesidir.
2. Kota Örneklemesi (Quota Sampling): Nitelik bakımından tabakalara ayrılmış olan kitleden her tabakayı temsil edecek kadar sayı belirlenir. Ve her tabakadan belirlenen kota kadar gözlem alınır.
3. Kartopu Örneklemesi (Snowball Sampling): Kartopu örneklemesi, araştırmacıların alkolizm, depresif ilaç kullanımı, seks işçileri, aile içi şiddet vb. gibi konularda kendini gizleyen ya da ulaşılması oldukça zor olan bireyler hakkında çalışma yaptıkları zaman kullandıkları olasılığa bağlı olmayan bir yöntemdir. Önce kitleden bir birime ulaşılır. Daha sonra birinci birim yardımıyla ikinci birime ulaşılır ve bu şekilde devam ederek birimlerden örneklem seçilir.
4. **Olasılığa Bağlı Örnekleme Yöntemleri**

Kitledeki her bir birimin örnekleme çekilme olasılığının (bu olasılık tam olarak belirlenebilir) sıfırdan farklı bilinen bir değere eşit olduğu örnekleme yöntemlerine, olasılığa bağlı örnekleme yöntemleri adı verilir. Bir başka deyişle olasılığa bağlı örnekleme yöntemlerinde, kitledeki hiçbir birimin örnekleme seçilme olasılığı sıfıra eşit değildir.

Olasılığa bağlı örnekleme yöntemleri, örneklem çekimi rastgele olarak yapıldığından rastgele örnekleme (random sampling) olarak da isimlendirilir. Örneklemin rastgele olarak çekilmesinden dolayı (her bir birimin örnekleme çekilme olasılığının eşit olduğu durumlarda), bu tür örnekleme yöntemlerinde örnekleme yanı (sampling bias) olmaz.

1. Basit Rasgele Örnekleme (Simple Random Sampling): Kitledeki birimlerin homojen olduğu durumlarda kullanılır. Her birime eşit seçilme olasılığı verilir. Bu yöntemde seçilen birimin yeniden seçilme olasılığı varsa eşit olasılıklı ya da yerine koyarak; seçilen birimin yeniden seçilme olasılığı yoksa eşit olmayan olasılıklı ya da yerine koymadan basit rasgele örnekleme olarak adlandırılır.

Rasgele sayılar tablosu kullanılarak, bilgisayardan rasgele sayılar üreterek ya da torbadan numara çekilerek seçme işlemi gerçekleştirilebilir.

1. Sistematik Örnekleme (Systematic Sampling): Kitle birimlerinin düzgün bir biçimde sıralanabildiği (özelliklerine göre) durumda ilk k (N/n) birimden herhangi biri başlangıç olarak seçilir ve daha sonra her k. birim seçilerek örneklem oluşturulur. Bu yöntem kolay olduğu için tercih edilir. Maliyeti düşüktür.
2. Tabakalı Örnekleme (Stratified Sampling): Kitle homojen yapıda değilse ve kendi içinde homojen alt birimlere (tabakalara) ayrılabiliyorsa kullanılan örnekleme türüdür. Örneklem her tabakadan ayrı ayrı ve birbirinden bağımsız seçilir. Seçilen örneklem sayısını (büyüklüğünü) tabaka büyüklüğüne göre belirleriz.
3. Küme Örneklemesi (Cluster Sampling): Örnekleme birimi bir demet, grup ya da küme şeklinde olabilir. Bu durumda kullanılan örnekleme türüdür. Örneğin; bir bankada çalışanların memnuniyeti için yapılan çalışmada her bir banka şubesi ayrı küme olarak değerlendirilip, her kümeden rasgele örneklem seçilir.

# SIKLIK ÇİZELGELERİ

**Diagram

Description automatically generatedDiagram

Description automatically generated**

**NİCEL VERİLER İÇİN SIKLIK ÇİZELGELERİ**

**Dağılım Genişliği (Range):** Verilerdeki maksimum ve minimum değer arasındaki farktır.

**Sınıf (Class)(k):** Verilerin ayrıldığı grup sayısıdır.

**k=1+3.3log (n)** ya da **√n** hesaplanır. n veri(denek) sayısını gösterir.

**Alt Sınır (AS):** Bir sınıfta yer alan en küçük değerdir.

**Üst Sınır (ÜS):** Bir sınıfta yer alan en büyük değerdir.

**Sınıf Aralığı (c):** Art arda gelen iki sınıfın alt ve üst değerleri arasındaki farktır.

A picture containing text

Description automatically generated

a değerleri tam sayı için a=1, virgülden sonra bir haneli olan ondalık sayı ise a=0.1 ya da virgülden sonra iki haneli olan ondalık sayı ise a=0.01 olur.

**Sıklık (Frekans)(fi):** Bir sınıfta bulunan gözlem sayısıdır.

f1+f2+f3+…+fk=n olmalıdır.

**Çetele:** Gözlemlerin düştüğü sınıfları belirlemek için yapılan işaretlemelerdir.

**Göreli Frekans (pi):** Her sınıfa düşen sıklığın toplam denek sayısına oranıdır. pi=fi/n

**Birikimli Sıklık (Fi):** Bir sınıfın sıklığı ve önceki sınıfların sıklıklarının toplamıdır. % Fi ise birikimli yüzdedir ve %Fi=Fi/n ile hesaplanır.

**Sınıf Orta Değeri (Si):** Her sınıfın alt sınır ve üst sınır ortalamasıdır.

Bir sıklık çizelgesinin sağlaması gereken özellikler:

* Veri kümesindeki en büyük ve en küçük gözlemi kapsamalıdır.
* Açıkta gözlem kalmamalıdır.
* k ve n orantılı olmalıdır.
* Sıfır sıklık olmamalıdır.
* Yoğunluk sınıfın ortalarında olmalı, uçlara doğru azalmalıdır.

**NİTEL VERİLER İÇİN SIKLIK ÇİZELGELERİ**

* Sıralanabilir nitel veriler için **fi, pi, Fi** ve **%Fi** değerleri çizelgede yer alır.
* Sınıflanabilir nitel veriler için **fi** ve **pi** değerleri çizelgede yer alır.

**EŞİT ARALIKLI OLMAYAN VE AÇIK UÇLU SIKLIK ÇİZELGELERİ**

* Bir veri kümesinde bazı sınıflara veri düşmemesi sonucu sınıfların birleştirilmesi ile ortaya çıkan sıklık çizelgeleri eşit aralıklı olmayan sıklık çizelgeleridir. Dağılım aşırı derecede çarpıklaştığında bazı sınıflardaki ayrıntıların kaybolup, bazı sınıftaki gereksiz ayrıntıların göründüğü durumlarda da kullanılabilir.
* Veri kümesindeki en küçük ya da en büyük değerin bilinmediği durumlarda açık uçlu sıklık çizelgeleri kullanılır.

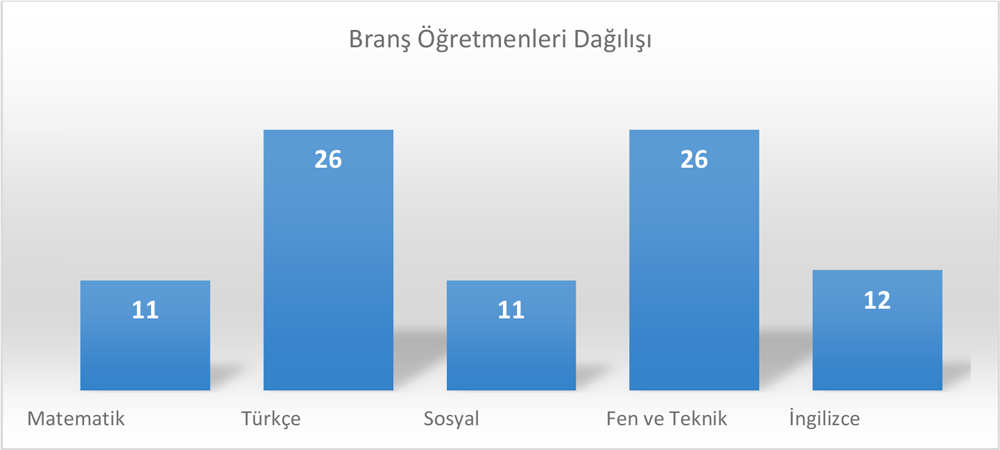
# GRAFİKLER

Grafik çizerken dikkat edilmesi gereken hususlar:

* Veri tipine uygun grafik seçilmelidir
* Çizilen grafiği açıklayıcı bir şekil numarası ve açıklaması verilmelidir.
* Eksenlerin neyi ifade ettiği gösterilmelidir.
* Ölçek belirtilmelidir.

1. **ÇUBUK GRAFİĞİ (BAR CHART)**

Nitel veriler için kullanılan grafik türüdür. Yatay eksende sınıflar yer alırken, dikey eksende sıklık/göreli sıklıklar yer alır.



1. **PASTA GRAFİĞİ (PIE GRAPH)**

Nitel veriler için kullanılan grafik türüdür. Sınıflandırılmış nicel verilerde de kullanılabilir. Her bir dilimin payını hesaplamak için Pi\*360 kullanılabilir.



1. **SİMİT (DONUT) GRAFİĞİ**

Nitel veriler için kullanılır.

Chart, sunburst chart

Description automatically generated

1. **PİRAMİT GRAFİĞİ**

Nitel veriler için kullanılır.

Shape

Description automatically generated

1. **KARE GRAFİĞİ**

Nitel veriler için kullanılır.

Chart, treemap chart

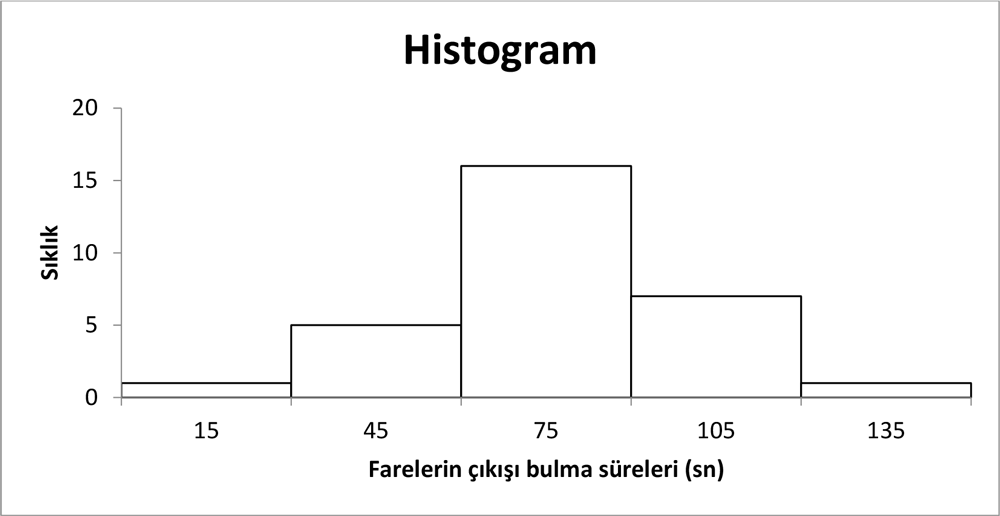
Description automatically generated

1. **HİSTOGRAM GRAFİĞİ**

Nicel (sürekli) veriler için sıklıkla kullanılan grafik türüdür. Çubuk grafiği ile benzerlik gösterir. Ancak histogramda çubuklar birbirine değmez. Si, fi ya da pi’leri gösterir. Sıklık çizelgesi oluşturulan bir veri seti için histogramdaki her bir çubuk ilgili sınıfın alt sınır değeri solda, üst sınır değeri sağda olacak şekilde işaretlemiştir. Sınıf alt ve üst değerlerini kullanmaktan çok yatay eksende ortalanmış biçimde sınıf orta değerlerini (Si) kullanmak (ilgili çubuğun altında) daha yaygındır.

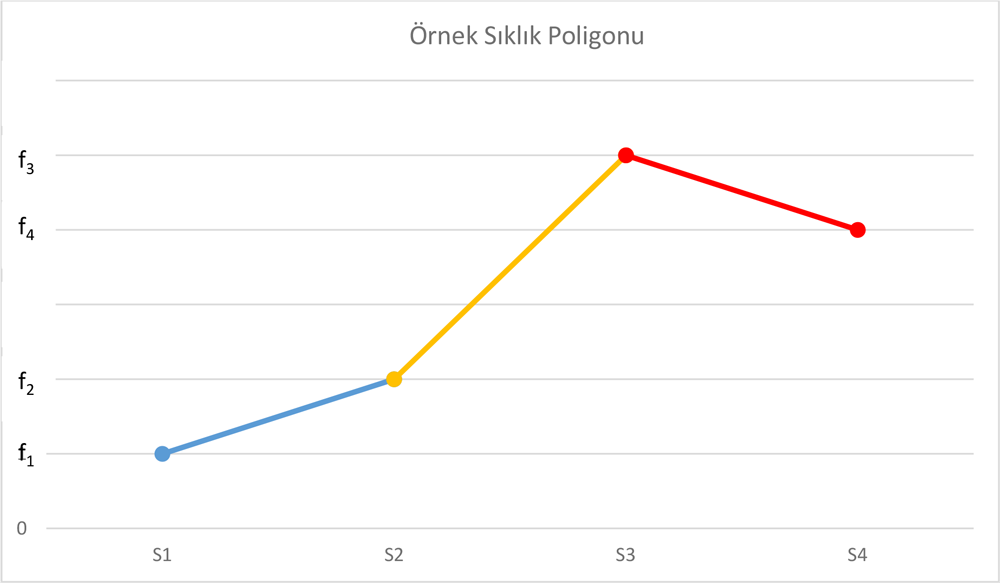
Histogram grafikleri:

* Verilerin merkezinin yaklaşık olarak nerede olduğunu,
* Verilerin yayılımını,
* Dağılımın şeklini (simetrik, çarpık vs.)
* Aykırı olan değerlerin olup olmadığını gösterir.



1. **POLİGON GRAFİĞİ**

Nicel (sürekli) veriler için kullanılır. Histogram grafiğinde tepe noktalarının doğru parçaları ile birleştirilmesi sonucu oluşur.



1. **BİRİKİMLİ SIKLIK POLİGONU (OGIVE)**

Nicel (sürekli) verilerde kullanılır. Yatay eksen ilk sınıfın alt sınır değeri ile başlar, son sınıfın üst sınır değeri ile biter. Dikey eksende Fi (birikimli sıklık değeri) değerleri bulunur.



1. **EĞRİ GRAFİĞİ**

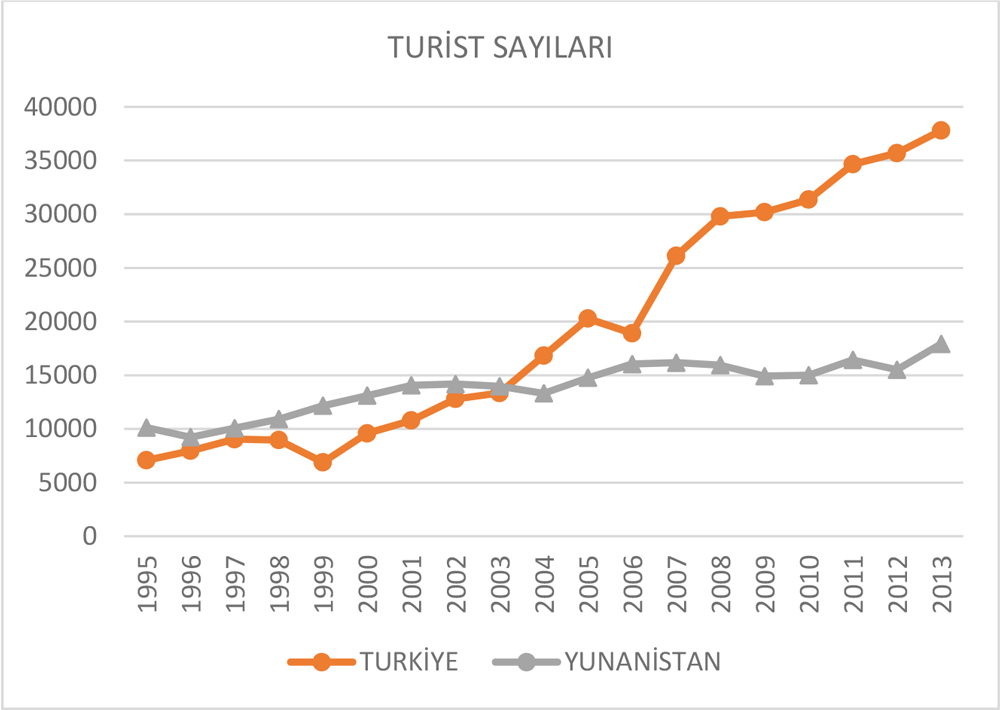
Sürekli nicel veriler için kullanılır. Poligon grafiğinin daha düzleştirilmiş şeklidir.

Chart, line chart

Description automatically generated

1. **ZAMAN SERİSİ (TREND) GRAFİĞİ**

Zamanın belirli periyotlarında (saatlik, günlük, haftalık, aylık, mevsimlik, yıllık, 5 yılda bir gibi) elde edilen verilerin zamana karşı eğilimini görmek için kullanılır. Bu sayede zaman serisi verilerinin zamana karşı kararlılığını, tepkisini ve şiddetini yorumlamak mümkündür. X ekseninde zaman, y ekseninde zamana karşılık gelen değer yer alır.



1. **DAL YAPRAK GRAFİĞİ**

Nicel verilerde kullanılan grafik türüdür. Veri sayısının az olduğu durumlarda kullanışlıdır. Verinin tümünü içerdiği için veri kaybı görülmez. Verinin merkezi ve yayılımı hakkında bilgi verir.

Veri değerlerinin basamak sayısına göre birler basamağı “yaprak”, geriye kalan basamaklar verilerin değerlerine göre “dal” olarak nitelendirilir.

Table

Description automatically generatedA picture containing shape

Description automatically generated

# KONUM ÖLÇÜLERİ

Konum ölçüleri verilerin dağılımlarını, birbirlerine göre konumlarını belirlemek için kullanılan ölçülerdir.

**NİCEL VERİLERDE KONUM ÖLÇÜLERİ**

Chart, diagram

Description automatically generated

1. **Aritmetik Ortalama**

**Diagram

Description automatically generated**

* Dağılımda bulunan tüm değerler dikkate alınır.
* Her bir verinin ortalamadan farklarının toplamı sıfırdır. ( ∑ (xi – ortalama)=0 )
* Dağılım sınırlarından (uç değerlerden) aritmetik ortalama etkilenir. Böyle değerlere sahip olunduğunda ortanca, tepe değeri gibi ölçüleri kullanmak daha yararlı olabilir.

A picture containing table

Description automatically generated

Sınıflandırılmış verilerde aritmetik ortalama bulmak için farklı bir yöntem.