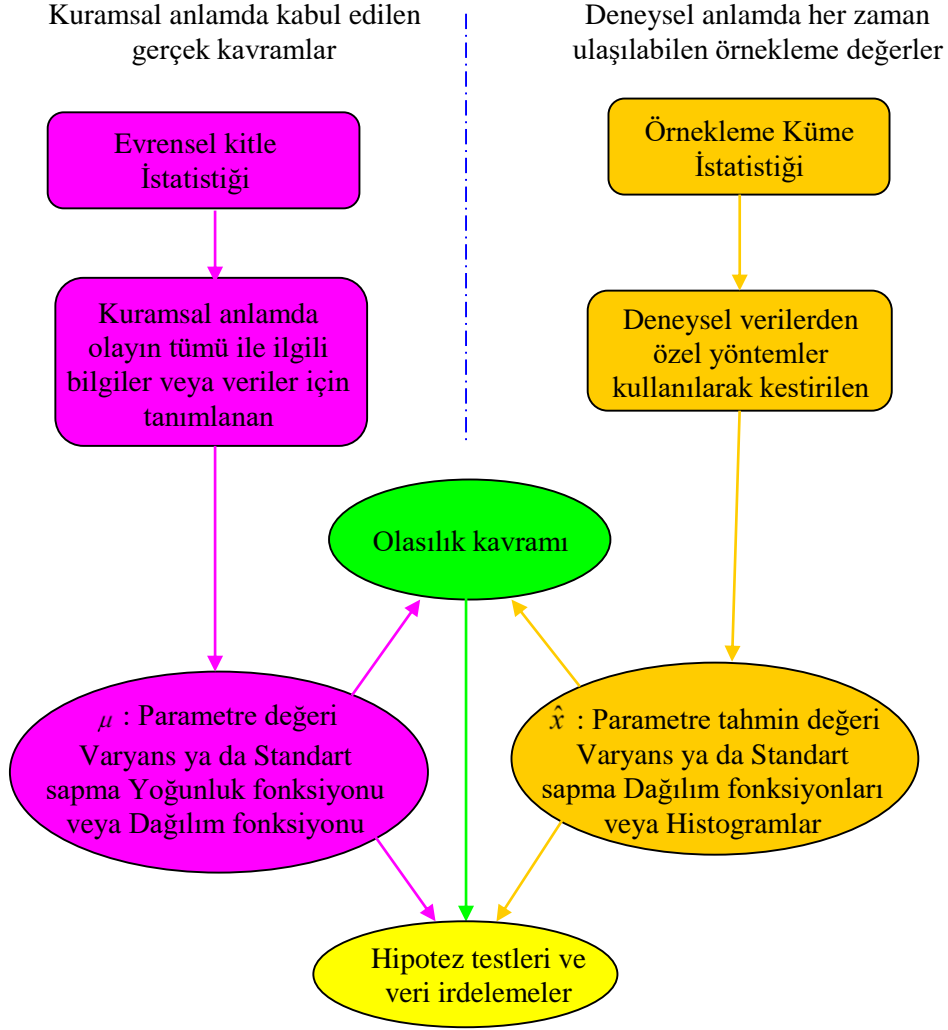


## İSTATİSTİK TEMEL BİLGİLER

### 3. İSTATİSTİK BAZI TEMEL KAVRAM VE TANIMLAR

Bilindiği gibi, istatistik çok geniş anlamda kullanılan bir sözcüktür. Bir yönüyle bu sözcük; herhangi bir olayla ilgili bilgilerin çeşitli şekillerde toplanması, depolanması, sonuç bilgileri üretilmesi, bu bilgilerin sergilenmesi, farklı şekillerde sınıflandırılmaları, çeşitli yöntemler kullanılarak değerlendirilmeleri, değerlendirme sonucunda elde edilen sonuç bilgilerinin sunulması, çeşitli şekillerde değerlendirilerek analiz edilmeleri...vs. gibi işlemlerin tümü için kullanılır. Böyle bir kullanıma göre bir örnekleme olayla ilgili tüm istatistik kavramlar, *Birinci adım istatistik*, *Olasılık İstatistik* ve *İkinci adım istatistik* şeklinde üç grup altında ele alınabilir. Bunlardan birinci adım istatistikte; bir olayla ilgili bilgilerin çeşitli şekillerde ve yöntemlerle toplanması, sonuçları, sergilenmeleri, sınıflandırılmaları, basit yöntemlerle değerlendirilmeleri ele alınmaktadır. İkinci adım istatistik olarak da, bilgilerin olasılık yasalarına göre farklı modellerle değerlendirilmeleri sonucunda elde edilen yeni istatistik sonuç bilgilerinin aynı şekilde değerlendirilmeleri konu edilmektedir. Olasılık istatistik ise; istatistiksel bilgilerin bir arada olasılık ya da ihtimaller kuramına göre incelenmeleri konu edilir. Bunların her biri, birbiriyle yakın ilişkisi olan iç içe girmiş kavramlar olmaktadır. Çoğu zaman kesin sınırlarla ayırt edilmeleri olanaksızdır.

Bu kısa açıklamadan sonra istatistik; özet olarak bir olayla ilgili gözlem ya da ölçü veya daha genel bir ifade ile denemeler sonucunda elde edilmiş verilerin toplanması, düzenlenmesi, özetlenmesi ve sergilenmesi, sunulması, bunların çeşitli olasılık-istatistik yöntemlerle analiz edilmeleri, elde edilen sonuçların grafikler, tablo ya da histogramlarla sunulması, aynı zamanda ilk verilerden veya çeşitli değerlendirmeler neticesinde elde edilmiş sonuçlara göre gerekli istatistik yorumların yapılarak geçerli kararların verilmesiyle ilgili bilimsel bir yöntemdir diye tanımlanabilir. Buradan görüldüğü gibi bu tanım pratik anlamda olup daha çok ulaşılabilen sonlu sayıdaki örnek veri kümeleriyle ilgili kavramları içermektedir. Kuramsal ya da evrensel kitle istatistiği ile fazla bir şey söylememektedir. Bu nedenle konu daha genel anlamda ele alınarak istatistik kavramlar ile ilgili Diyagram 1'deki gibi bir özetleme veya tanım yapılabilir.



*Diyagram 1: İstatistik kavramlara genel bir bakış diyagramı*

Diyagram 1 de özet biçimde açıklandığı gibi daha genel anlamda; gerek bir olayla ilgili sonlu sayıda veriden oluşan bir örnekleme küme, gerekse olabilecek tüm durumları ifade eden evrensel kitlelerin temel istatistikleri ile ilgili istatistik kavramlar en genel şekliyle iki farklı yaklaşımla ele alınabilir. Bunlardan birincisi; bir olayla ilgili olarak tüm olabilecek sonuçları içeren evrensel küme istatistiğini, bir diğeri de; evrensel küme elemanlarının sonlu sayıda örnekleme sonuçlarını içeren ve her zaman ulaşılabilen durumdaki örnekleme küme istatistiğidir.

Ancak burada özetle söylemek gerekirse; Diyagram 1'de genel anlamda kendi doğal ortamları içerisinde verilmiş olan tüm konular, daha ayrıntılı bir biçimde ele alınarak özel yöntemler kullanmak suretiyle daha detaylı bir şekilde ayrıca değerlendirilebilir. Bu amaçla, Diyagram 1'de özet olarak verilmiş olan bazı istatistik temel kavramlar; böyle bir genel diyagram mantığı içerisinde ele alınarak, her bir olayla ilgili kullanılan bazı özel istatistik temel terim ve tanımlar aşağıdaki gibi açıklanabilir.

### 3.1. İstatistiksel Veri (*Data*)

Doğal olay ya da denemelerin gözlenmeleri veya ölçülmeleri sonucunda elde edilen analitik değerlerdir. Her zaman bunlar bir reel sayı özelliğinde olurlar. Ayrıca bu değerler, özellikleri gereği, herhangi bir istatistikte farklı istatistik yöntemlerin uygulanmasından elde edilecek farklı bilgilerin kesin bir dille ifadesi olan sayısal ya da nicel özelliklere sahip haber vericilerdir denebilir. Bu nedenle, daima tek anlam taşırlar.

### 3.2. İstatistik Değişken

Bir olayla ilgili herhangi bir istatistik örnekleme sonucunda elde edilen özel değerlerinden herhangi birini özetleyebilen değişiklilerin en genel gösterime sahip nicel bir ifadesidir. Bu amaca yönelik tanımlanan tüm değişkenler, sürekli ya da süreksiz türden ifadeler olabilirler. Buna göre;

- **Sürekli değişken:** Belli bir aralıkta olabilecek tüm değerleri ifade edebilen ya da özetleyebilen bir değişken olarak tanımlanırlar.
- **Ayrık değişken de:** Herhangi bir aralıkta belli değerleri özetleyebilen diskirit (*ayrık*) türden özelliğe sahip bir değişken olduğu şeklinde tanımlanabilir.
- **Bir sabit:** Belli bir aralıkta sadece belli değeri ifade edebilen ayrık türden bir değişken olmaktadır. Belli aralıkta daima tek değere sahiptir.

Neticede, jeodezik ölçme işlemi sürekli türden değişken, buna rağmen her bir ölçmeye karşılık gelen sayısal sonuçlar da ayrık türden değişken olmaktadır.

### 3.3. Rastgele Değişken

Bu tanımların paralelinde bir rastgele değişken; kendine özgü rölatif sıklıkta (*frequency*) veya olasılıkta (*probability*) *spesifik* bir örnekleme değerlerinden belli sınırlar arasındaki herhangi birini özetleyebilen kantitatif özelliklere sahip bir değişken olarak tanımlanabilir. O zaman istatistik olayları tanımlayan bir rastgele değişken sadece sıradan değerlerin oluşturduğu bir veri örnekleme tarafından değil de aynı zamanda onların bir ortalama değer etrafındaki dispersiyonunu da ifade eden, nasıl bir sıklıkta ya da *frekansta* oluştuğunu gösteren, *yoğunluk* ya da *dağılım* fonksiyonları ile birlikte ancak tanımlı kılınabilirler.

### 3.4. Parametre

Bir olayla ilgili ana kitleyi tanımlayan ortak sayısal belirleyiciler olarak tanımlanabilir. Matematik-istatistikte bunlar çeşitli olabilirler. Ana kitleye ilişkin  $\mu$  ortalama ve  $\sigma^2$  varyans değerleri bu amaçla kullanılmakta olan ana kitle karakteristiklerinden biri olmaktadır.

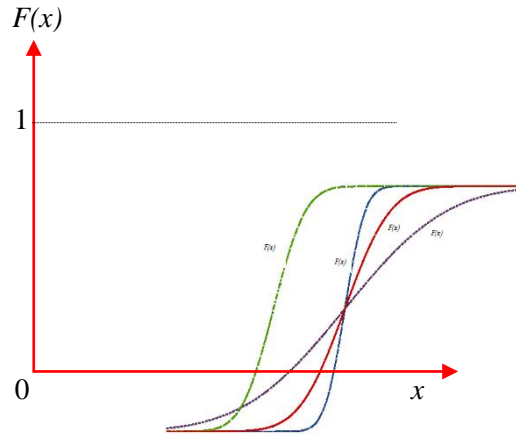
### 3.5. Evrensel Küme ya da Kitle (*population*)

Bir yığın veya diğer bir ifade şekliyle evrensel küme ya da kitle (*population*); “özel ölçülebilir bir rastgele değişken tarafından ifade edilen bütün nesnelere bir koleksiyonudur” şeklinde tanımlanabilir. Böyle bir tanım aynı zamanda bütün ve ortak özellikleri olan canlı ya da cansız her türlü elemanları içerebilir. Aynı zamanda evren sözcüğü, tekli elemanlar için “*örnek olay*”, küçük çokluklar için “*araştırma kümesi*” gibi deyimlerle de ifade edilebilir. Sonuçta, bu haliyle basit bir olay için tek birimli olabilecekleri gibi çok birimli de olabilirler. Genelde, tanımlanmış oldukları olayların özellikleri gereği, bir kitle ya da evrensel küme sonlu veya sonsuz sayıda elemanlardan oluşabilir. Bu nedenle, bir olayla ilgili olabilecek bütün sonuçları içerirler.

Bunun neticesinde, evrensel kümenin istatistik tahmin parametreleri daima kuramsal anlamdaki dağılım parametreleri olmaktadır. Örneğin; bir paranın sadece bir kez yazı tura atılması ile elde edilen olası sonuçların oluşturduğu yığın ya da evrensel küme (*Population*); yazı ya da tura gibi iki elemandan ibarettir ve aynı zamanda sonlu sayıdadır. Buna karşılık, hem bir paranın aralıksız çok sayıda yazı tura atılması ile elde edilen olası sonuçlarından oluşan bir yığın veya evrensel küme sonsuz sayıda olmakta; hem de 0 ve 1 arasındaki bütün gerçek sayılardan oluşan yığın ya da evrensel küme aynı şekilde sonsuz sayıda olmaktadır. Her durumda, bu olaylarla ilgili her bir eleman aynı zamanda evrensel kümelerin tek bir üyesi olur.

### 3.6. Dağılım Fonksiyonu

Belli sınırlar arasında, bir rastgele değişkene rastlama olasılığıdır. Uygulamada, böyle bir fonksiyon aynı rastgele değişkenin hangi sıklıkta meydana geldiğini ifade eden, yoğunluk fonksiyonunun integrali biçiminde de elde edilebilir. Böyle bir fonksiyona bir örnek, farklı parametre değerlerine göre verilmiş olan her bir normal dağılımla ilgili  $f(x)$  yoğunluk fonksiyonlarına karşılık gelen farklı parametre değerlerine sahip  $F(x)$  normal dağılım fonksiyonları için, ilgili dağılım eğrileri,

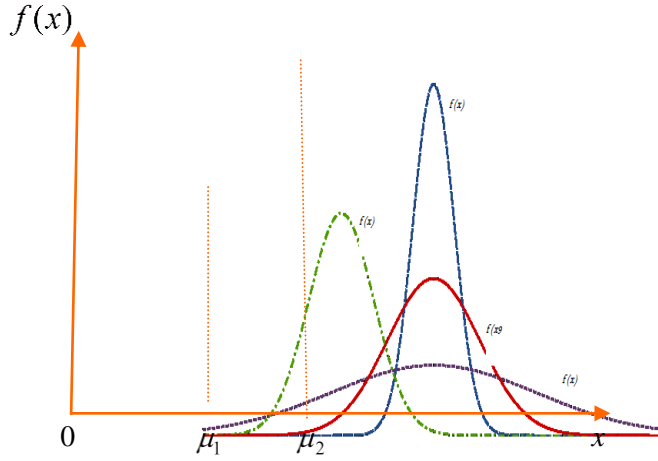


Şekil 3: Farklı parametrelili  $F(x)$  dağılım eğrileri

şeklinde verilebilir (Şekil 3).

### 3.7. Yoğunluk ya da Frekans Fonksiyonu

Belli sınırlar arasında, bir rastgele değişkenin alacağı değerlerin yayılma alanı içinde  $\mu$  ortalama etrafındaki dispersiyon sıklığını ya da yoğunlaşmasını ifade eden bir fonksiyondur. Bunlar üstel yapıya sahip doğal fonksiyonlar olup dağılım fonksiyonunun türevi şeklinde de bulunabilirler. Bazı uygulamalarda olasılık fonksiyonu olarak da adlandırılmaktadırlar. Yoğunlukla üstel fonksiyonlarla ifade edilirler. Konuya bir örnek olarak farklı parametre değerlerine göre çizilmiş normal dağılımla ilgili  $f(x)$  yoğunluk fonksiyonları eğrileri, Şekil 4 'deki biçimde verilebilir.



Şekil 4: Farklı parametrelili  $f(x)$  yoğunluk fonksiyonları

biçiminde verilebilir(Şekil 4).

### 3.8. Örneklem Küme veya Cümle

Bir olayla ilgili belli sayıda yapılan denemelerin sonuçlarını içeren bir kümedir. Böyle kümeler daima sonlu sayıda elemanlardan ya da verilerden oluşur. Bu nedenle, evrensel veri kümesinin bir alt kümesi olmaktadır. Bağımlı olduğu evrensel kümenin bütün özelliklerini taşır. Bu gibi özelliklerinden dolayı, her zaman evrensel kümenin tümü hakkında genel bir yargıya varmak için kullanılabilir. Örneklem küme bir örnek; bir derste tüm erkek öğrencilerin oluşturduğu küme verilebilir. Bu küme aynı zamanda bütün öğrencilerin oluşturduğu evrensel öğrenci kümesinin bir alt kümesi olmaktadır. Bu gibi özelliklere sahip bir örneklem küme için iki farklı ön koşulun sağlanmış olması gerekir. Bunlar; seçilen örneğin temsil yeteneği taşıması yani ana kitlenin özelliklerini tam olarak yansıtabilmesi ve örnek hacminin yeterli sayıda ya da büyüklükte olması şeklinde sıralanabilir. Bir diğer özelliği ile örneklem kümesinin ya da alt grup kümesinin oluşturulmasında kullanılan yöntemler *rastgele* örneklemeler olabilecekleri gibi yargısal özellikteki örneklemeler de olabilirler. Tesadüfi yani rastgele özellikte olan örneklemelerde, örnek seçimi sadece olasılık kurallarına dayanmaktadır. Bu nedenle, rastgele örneklemeyi de kendi içerisinde *basit* rastgele örneklem ve *tabakalı* örneklem olmak üzere iki farklı yaklaşıma ayırmak mümkün olmaktadır. Bilindiği gibi, basit tesadüfi örneklemde ana kitleyi oluşturan her elemanın seçilme şansı daima eşittir. Halbuki, tabakalı örneklem yönteminde, anakitleyi çeşitli alt bölümlere ayırarak her bir alt bölümden rastgele bir yaklaşımla seçimler yapılır. Buna karşılık, tabakaların kendi içinde her zaman türdeş olmaları büyük önem taşımaktadır. Sonuçta, tabakalı

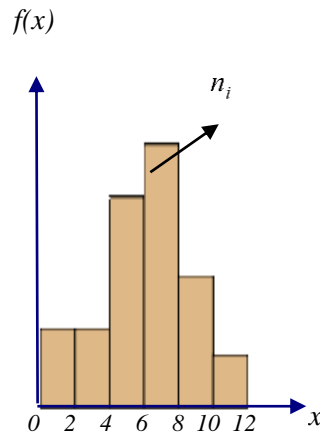
örneklemenin yüksek temsil olanağı sahip olduğu söylenebilir. Bir diğer yönüyle, yargısal örneklemelerde rastgelelik ilkesine göre hareket edilmeden elde edilen verilere göre ana kitleyi en iyi temsil edeceğine karar verilen bir alt grup örnek olarak alınır. Neticede, bunun üzerinden yapılan gözlemler, irdemeler sonucu elde edilen sonuçlar ana kitleye genelleştirilir.

### 3.9. Sınıf, Sınıf Sınırları, Sınıf Aralığı

Bir örneklemedeki elemanlar uygun değişken aralıklarına bölünmesine göre gruplara ya da sınıflara ayrılabilirler (Şekil 5). Bu şekilde tanımlanan her bir grup bir sınıf olarak adlandırılır. Bir sınıfın üst ve alt sınırları, sınıf sınırları olarak isimlendirilir. Bu sınıf sınırları arasındaki aralık sınıf aralığı olarak adlandırılır.

### 3.10. Sınıf Yığılmaları veya Frekansları

Belli bir sınıfa düşen elemanların sayısı sınıf yığılmaları veya frekansları olarak adlandırılır. Rölatif yığılma (kısmi yığılma olarak bilinir) örneklemedeki toplam eleman sayısının bir bölümü olarak ifade edilen sınıf yığılmasıdır. Aynı zamanda bu değer  $n$  sayıda elemandan oluşan bir veri kümesinde bir sınıfa düşen eleman sayısı  $n_i$  ise, bunun bağıl sınıf yığılmaları,  $p_i = \frac{n_i}{n}$  olarak ifade edilebilir. Böyle bir değer aynı zamanda sınıf frekansları olarak da adlandırılmaktadır.



Şekil 5: Histogram dağılımı

Uygulamada, bir olayla ilgili tüm örnekleme sonuçları için bu şekildeki gösterimler bir arada ele alındığında örnekleme verilerin *histogram* dağılımı ifade edilmiş olur. Böyle bir histograma herhangi bir olayla ilgili veri kümesinden belli sınıf aralıklarına göre çizili histogram dağılımı grafiğini göstermektedir (Şekil 5).

### 3.11. Olasılık Kavramı

Bugüne kadar geçen sürede, olasılık kavramı ile ilgili, istatistikle uğraşan bilim adamları tarafından birçok tanım verilmiştir. Bunlardan birincisi her yerde sıkça rastlanan olasılığın klasik tanımıdır. Olasılık istatistiğinin klasik tanımı; bir yığından (*population*) eşit ihtimalle seçilen  $n$  sayıda elemanın, yığındaki  $N$  tüm eleman sayısına oranı

$$P_r(A) = \frac{n}{N}$$

olarak tanımlanmaktadır. Daha kısa başka bir ifade ile olasılıkla ilgili bu tanım; bir olayla ilgili sonuçlardan uygun haller sayısının mümkün haller sayısına oranıdır şeklinde de söylemek mümkündür. Burada, olasılığı tanımlamak için kullanılan eşit ihtimal kelimesi bu gün için yuvarlak bir ifade olup eşit koşullar anlamında kullanılmaktadır (*Wells, D. E.–Kraikowsky 1971*).

Uygulamada böyle bir olasılık sorunu tamamıyla birbiriyle aynı olmayan iki şekilde çözülebilir. Bunlardan birincisi; bir olayla ilgili elde edilen  $N$  adet deneysel verilerden,  $k$  tane sınıftan biri olan, özel bir  $A$  sınıfına düşecek  $n$  elemanlar sayısından,

$$P_r(A) = \frac{n}{N}$$

ilişkisine göre *ampirik* olarak belirlemektir. Böyle bir yaklaşım inanış biçimine dayalı bir tanım olmadığından analitik bir anlam taşımaktadır. Bu nedenle, pratikte oldukça yaygın bir biçimde kullanılmaktadır.

Olasılık istatistiğinin diğer bir tanım şekli de inanış biçimine göre yapılan tanımdır. “*Bu duruma bir örnek; uzaktan gelen bir insanı görünce onun hemen iyi bir insan olduğunu söylemek gibi bir karar; ikinci tür olasılık tanımına bir örnek verilebilir. Bu karar tamamıyla sezgi biçimine bağlıdır.*” Bu ikinci tanım; aksiyonlara göre olasılık kurallarını ifade eden veya sezgi biçimine göre karar verilen tanımsız bir kavramdır (*Papoulis 1965*).



Bu nedenle, böyle bir tanım tamamıyla sezgi biçimine dayanmaktadır. İnanış biçimine göre yapılmış olan bu ikinci tanım; özelliği gereği, ancak *kantitatif* olmayan olayların incelenmesinde kullanılabilir bir tanım ve yöntem olmamaktadır. Buna karşılık, *kantitatif* olayların olasılık yönünden irdelenmesinde, genellikle, parametrik olmayan bu ikinci tanım yerine, birinci tanım ya da bu tanımdan elde edilen kurallar veya sonuçlar kullanılmaktadır.

Sonuçta burada konunun tekrar bir özeti gerekirse; *Diyagram 1* 'de de genel anlamda görüldüğü veya söylendiği gibi, bir olayla ilgili bilgilerin tümü, diğer bir ifade ile “*bir rastgele değişkenin alabileceği değerlerin tümü*” bir evrensel kümede ya da kitlede ele alınarak küme ile ilgili evrensel kitle istatistiği ve ilgili dağılım parametreleri kuramsal anlamda tanımlanabilir. Daha sonra bu gibi bilgiler, kuramsal istatistiğin temel kavramlarını oluşturacağı için, bunlarla ilgili tüm matematiksel ilişkiler benzer matematiksel varsayımlara göre geliştirilir. Aynı zaman bu bilgiler kuramsal istatistik temel kavramlarını olmaktadır.

Ne var ki, pratikte ya da deneysel çalışmalar sonucunda böyle bir evrensel kitle kümesine her zaman ulaşmak mümkün olmamaktadır. Bunun yerine, benzer istatistikler sonlu sayıda örnekleme elemanları içeren örnekleme veri küme ya da cümleleriyle yürütülür. Diğer bir ifade ile bir rastgele değişkenin belli aralıkta alacağı belli değerlerden oluşan bir örnekleme küme istatistiği ile gerçekleştirilir. O zaman kuramsal istatistik parametreleri yerine sadece deneysel verilere göre tahmin edilmiş kestirim parametre değerleri kullanılır ve aynı zamanda bilinebilir.

Örnekleme sonucunda sonlu sayıda veriden oluşan bu örnekleme küme istatistiği bilgi ve değerlerinin gerçeği ya da evrensel küme bilgilerini ne derece yansıttığını söyleyebilmek, aynı zamanda bu örnekleme küme bilgilerinin veya onlardan üretilecek sonuç değerlerinin gerçek veya evrensel küme istatistiği sonuçlarını ne derece yansıttığını, bunlara dayalı yorumların ne derece inandırıcı olduklarını açıklayabilmek için haklarında istatistik anlamda çeşitli hipotezler kurularak belli anlamlılık seviyesine göre matematik istatistik yöntemlerle irdelenir. Uygulamada böyle bir konu hipotez testi veya veri irdemesi olarak adlandırılmaktadır.

Jeodezi'de böyle bir konu, ilk bağımsız gözlemlerden kurulu örnekleme veri kümelerine göre yapılabileceği gibi, uygun istatistik parametre kestirim yöntemleri kullanılarak bunlardan sonuç ürünü olarak tahmin edilmiş diğer bir ifade ile üretilmiş veri kümesi değerlerine göre gerçekleştirilir. Neticede konu, ölçme ya da gözlemler sonucunda elde edilmiş sonlu sayıdaki jeodezik veriden oluşmuş örnekleme veri kümelerinden; doğada her zaman var olan gerçek bir olayla ilgili durumu temsil eden evrensel küme hakkında karar ve yorum yapma ilkesine dönüşmüş olur. Burada hatırlanacağı gibi; pratikte bu olay jeodezik verilerin irdelenmesi olarak adlandırılır.

Ancak, burada tekrar vurgulamak gerekirse; böyle bir irdelemenin doğru ve inandırıcılığı, her şeyden önce örnekleme veri kümesinin evrensel kümeyi alabildiğince en iyi şekilde temsil etmesine bağlıdır. Bu amaçla, her bir olayla ilgili bütün örnekleme veri kümelerinin elde edilmesinde en başta dikkat edilmesi gereken temel kural; bütün ölçü ya da gözlem değerlerinin tamamıyla yansız, daha açık bir ifade ile hiçbir kaba ve sistematik olumsuz etki içermeyecek şekilde, aynı zamanda evrensel veri kümesini en iyi şekilde temsil edecek yeteri sayıda elde edilmiş rastgele veriden oluşturulmuş olmalarıdır. Bu gibi özellikleri içermeyen örnekleme veri kümeleriyle verilecek kararlar aynı şekilde yanıltıcı olduğu kadar yanlış yorumlara da neden olur.

Bu amaçla burada konu; önce evrensel küme istatistiği biçiminde ele alınıp genel durumuyla kuramsal olarak özet bir biçimde açıklandıktan sonra devam eden diğer paragraf ve bölümlerde örnekleme küme istatistiği ile ilgili işlemler şeklinde verilmektedir. Ancak, burada konu jeodezik verilerin irdelenmesi biçiminde düşünülmüş olduğundan, örnekleme veri kümesi de; genel veri kümesinden daha çok jeodezik verilerden oluşan bir örnekleme veri kümesi olarak ele alınıp, jeodezik uygulamalardaki özel durumu açıklanmıştır.