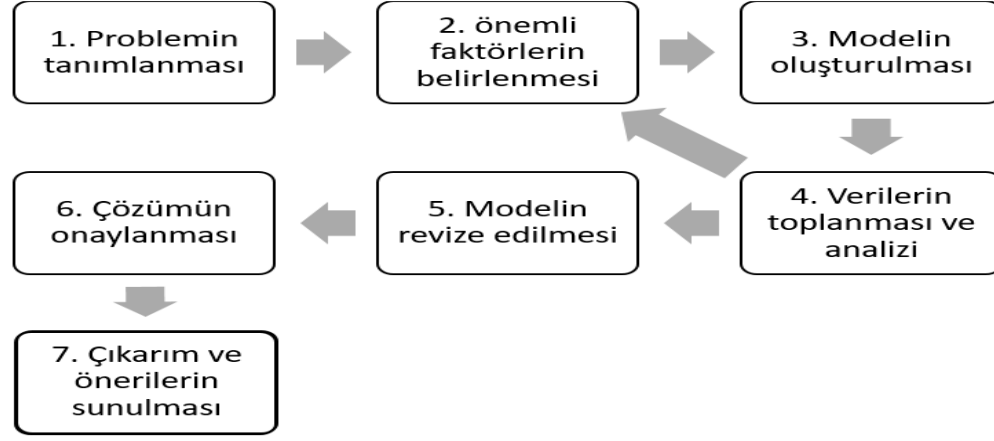


OLASILIK VE İSTATİSTİK DERS NOTU

1. TEMEL İSTATİSTİKSEL KAVRAMLAR

Problem çözümünde istatistiğin rolü:



Örnek: Araç yakıt tüketimini etkileyen faktörler ve etki büyüklükleri üzerine araştırma yapılsın. Motor hacmi, Motor gücü, Yüksek ve düşük devirde kullanım, Araç içi ağırlık, Klima kullanımı, Araç rengi, Araç donanımı gibi faktörleri ele alalım.

1.1. Tanımlamalar:

İstatistik bilimi; verilerin toplanması, düzenlenmesi, özetlenmesi, sunumu, analizi ve bu analizler aracılığıyla elde edilen sonuçların yorumlanması ve bir karara bağlanması ile ilgilenir. Çağdaş anlamda istatistik ise doğa olaylara dayanan ve gözlemlerle bulunan verinin bilimsel yöntemlerle incelenmesi ve doğru sonuç çıkarılması için kullanılan tekniklerin tamamıdır. Şu halde istatistiği ikiye ayırabiliriz:

- 1) Betimsel istatistik
- 2) Çıkarımsal istatistik.

Frekans dağılımları, merkezi ölçüleri (aritmetik ortalama, mod, medyan, ...), dağılış ölçüleri (standart sapma, varyans, değişim aralığı, ...), çarpıklık ve basıklık gibi konular verilerin özetlenmesi ve betimlenmesi ile ilgili olup betimsel istatistiğin konusudur. Örneklem teorisi, tahminleme, hipotez testleri, ilişki katsayıları ve regresyon analizi gibi konular ise çıkarımsal istatistiğin konusudur.

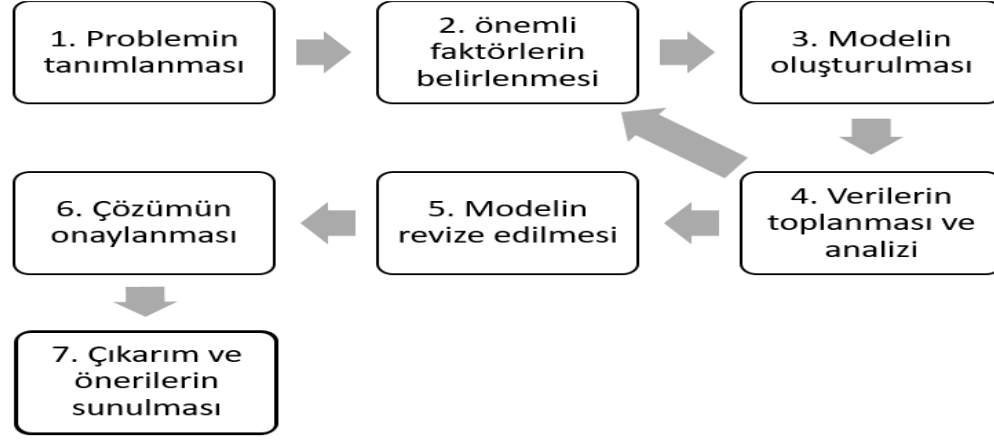
Kitle: Üzerinde ölçüm ve araştırmaların yapılacağı, belirli bir ya da birkaç özelliği incelenecek olan birimler (insanlar, nesneler, olaylar, kayıtlar) topluluğudur. Örneğin; nüfus sayımı için kitle Türkiye olup Denizli'deki üniversite öğrencilerinin aylık giderleri için kitle PAÜ öğrencileridir.

Tamsayım: Kitle hakkında tüm bilgilerin elde edilmesi kitleyi oluşturan tüm birimlerin incelenmesi demektir. Kitlenin tüm birimlerinin incelenmesi işlemine tam sayım denir.

OLASILIK VE İSTATİSTİK DERS NOTU

1. TEMEL İSTATİSTİKSEL KAVRAMLAR

Problem çözümünde istatistiğin rolü:



Örnek: Araç yakıt tüketimini etkileyen faktörler ve etki büyüklükleri üzerine araştırma yapılsın. Motor hacmi, Motor gücü, Yüksek ve düşük devirde kullanım, Araç içi ağırlık, Klima kullanımı, Araç rengi, Araç donanımı gibi faktörleri ele alalım.

1.1. Tanımlamalar:

İstatistik bilimi; verilerin toplanması, düzenlenmesi, özetlenmesi, sunumu, analizi ve bu analizler aracılığıyla elde edilen sonuçların yorumlanması ve bir karara bağlanması ile ilgilenir. Çağdaş anlamda istatistik ise doğa olaylara dayanan ve gözlemlerle bulunan verinin bilimsel yöntemlerle incelenmesi ve doğru sonuç çıkarılması için kullanılan tekniklerin tamamıdır. Şu halde istatistiği ikiye ayırabiliriz:

- 1) Betimsel istatistik
- 2) Çıkarımsal istatistik.

Frekans dağılımları, merkezi ölçüleri (aritmetik ortalama, mod, medyan, ...), dağılış ölçüleri (standart sapma, varyans, değişim aralığı, ...), çarpıklık ve basıklık gibi konular verilerin özetlenmesi ve betimlenmesi ile ilgili olup betimsel istatistiğin konusudur. Örneklem teorisi, tahminleme, hipotez testleri, ilişki katsayıları ve regresyon analizi gibi konular ise çıkarımsal istatistiğin konusudur.

Kitle: Üzerinde ölçüm ve araştırmaların yapılacağı, belirli bir ya da birkaç özelliği incelenecek olan birimler (insanlar, nesneler, olaylar, kayıtlar) topluluğudur. Örneğin; nüfus sayımı için kitle Türkiye olup Denizli'deki üniversite öğrencilerinin aylık giderleri için kitle PAÜ öğrencileridir.

Tamsayım: Kitle hakkında tüm bilgilerin elde edilmesi kitleyi oluşturan tüm birimlerin incelenmesi demektir. Kitlenin tüm birimlerinin incelenmesi işlemine tam sayım denir.

Karakteristik : Bir kitlenin her bir değişkeni için ölçülebilir olan değişkene, bu kitlenin karakteristiği denir.

Örneklem: Kitleye ulaşmanın mümkün olmadığı ya da yüksek maliyetler gerektirdiği durumlarda kitleden belirli yöntemlerle seçilen gözlemlerin oluşturduğu kümeye denir.

Örnekleme: Kitleden, kitle birim sayısından daha az sayıda birimin seçilmesi ile kütle hakkında tahmin yapma işlemlerine denir.

Bir deneyin uygulama aşamasında, araştırmacı karşılaştığı kitlenin tüm elemanlarını inceleyemeyeceği için örneklem üzerinde çalışmak kaçınılmaz olabilir. Örneğin bir hastalığa yönelik bir araştırmada ilgili bölgedeki bütün hastalar incelenemeyeceği için kitlenin sonlu sayıdaki elemanın karakteristiklerinin yani örneklemin incelenmesi ile sonuca ulaşılacaktır. Diğer bir taraftan ise örneklem, yukarıdaki gibi zaman, kaynak vs. açısından çalışmayı olanaklı hale getirmesi açısından seçilmenin yanında bir zorunluluk da olabilir. Örneğin tarih öncesi çağda yaşayan insanların kafatası büyüklüğünü ölçmek isteyen bir antropolog açık olarak bütün kitleyi deneye dahil edemez, örneklem ile çalışmak gerekli ve hatta zorunlu olacaktır.

Gözlem: Gözlemlenen veya kayıt altına alınan her bir ölçüme denir.

Değişken: Gözlemden gözleme değişik değerler alabilen objelere, özelliklere ya da durumlara denir. Örneğin; cinsiyet, eğitim düzeyi, boy, kilo, aylık harcama.

Bir örnek vermek gerekirse, Ankara’da yaşayan yetişkinlerin boy uzunlukları ile ilgili çalışılıyor ise bu durumda kitle Ankara’daki tüm yetişkinler, karakteristik her bir kişinin boy uzunluğu, parametre ise kitlenin ortalama boy uzunluğu olabilir.

Bir deneysel çalışmaya başlamadan önce kitlenin ve ölçülecek karakteristiklerin net bir şekilde tanımlanmış olması vazgeçilemez başlangıç şartıdır.

Parametre ve Örneklem İstatistiği: Kitlenin tanımlayıcı sayısal değerlerine parametre, örneklemin tanımlayıcı sayısal değerlerine de örneklem istatistiği denir. İstatistiksel çıkarımda amaç örneklem istatistiği kullanılarak kitle parametreleri hakkında çıkarımda bulunmaktır.

	Parametre	Örneklem istatistiği
Ortalama	μ	\bar{x}
Varyans	σ^2	s^2
Standart sapma	σ	s
Oran	Π	p
Gözlem sayısı	N	n

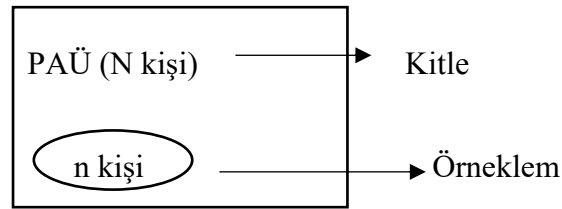
Örnek: PAÜ deki öğrencilerin aylık harcamaları hakkında araştırma yapılsın.

Kitle: PAÜ de kayıtlı öğrenci sayısı $N=46740$

Parametre (μ): öğrencilerin ortalama aylık harcaması

Örneklem: Rasgele seçilen $n=250$ öğrenci

Örneklem istatistiği (\bar{x}): örneklem ortalaması



Örneklem seçiminde dikkatli olmak gerekir. Kötü bir örneklem seçimi yapılmışsa matematiksel ve istatistiksel yöntemler bizi doğru sonuca götürmekten uzak olacaktır. Örneklem seçimi için en önemli nokta örneklem kitle için temsil edici olmasıdır. Ankara ilindeki yetişkinlerin boy uzunlukları ortalamasının hesaplanacağı bir çalışmada örneklem basketbol oyuncularından seçilirse, örneklem kitlenin özelliklerini temsil edemeyecektir.

Bir kitle sonlu ya da sayılabilir sayıdaki elemanlardan oluşmuşsa kitleye kesikli kitle, aksi halde sürekli kitle adı verilir. Bir A ilacının kullananların sayısı kesikli olmakla birlikte kullanım süresinin uzunluğu süreklidir.

1.2. Verilerin Toplanması

- **Geçmişe Dönük Araştırma:** Geçmişteki bir zaman diliminde toplamış verilerin bir kısmının ya da tamamının kullanılması durumudur. Maliyet ve zaman tasarrufu sağlamaktadır. Fakat verinin geçerliliği ve güvenilirliği şüphelidir. Zaman içerisinde gözlem kaybı yaşanmış olabilir. Veri toplanırken hatalar yapılmış olabilir. Tam amacımıza hitap etmiyor olabilir. Örnek: Eski laboratuvar çalışmasında elde edilen veriler.
- **Gözleme Dayalı Araştırma:** Bir süreç veya kitlenin belirli bir periyotta gerçekleştirilen rutin işlemler sırasında verilerin gözlemlenip kayıt altına alınmasıdır. Örnek: Yeni kurulan elektrik santralinin bölgedeki sıcaklık değişimine etkisi.
- **Deney Tasarımı:** Sistemde yer alan kontrol edilebilir değişkenler (faktörler) üzerinde bilinçli değişiklikler yapılarak sistemin bu değişikliklere verdiği tepkiyi gözlemleyerek değişkenlerin çıktı performansının etkilerini belirleyebilmesidir. Örnek: Araç yakıt tüketimini etkileyen faktörler örneği.
- **Bazı Örneklem Yöntemleri:**

- Basit rasgele örneklem: Bir kitleden her bir elemanın seçilme şansının eşit olduğu örnekleme yöntemidir. Örneğin, PAÜ aylık ortalama harcama örneği gibi rasgele 250 öğrenci seçilmesi.
- Sistematiik örneklem: Örnekleme oluşturacak ilk eleman kitleden rasgele seçilir, diğer elemanların ise belirli ve aynı kural çerçevesinde seçimi ile belirlenen yöntemdir. Genel olarak (örneklem aralığı=Kitle büyüklüğü/örneklem büyüklüğü) formülü ile rasgele seçilen ilk elemana örneklem aralığı eklenerek oluşturulur. Örneğin sınıftaki 80 öğrenciden 10 kişilik örneklem seçelim. Örneklem aralığı=80/10 yani 8 olup sınıf listesinden rasgele 2 öğrenci ilk eleman olarak seçilsin. O halde örnekleme 2, 10, 18, 26, 34, 42, 50, 58, 66 ve 74 sıra numarasına sahip öğrenciler oluşturacaktır.
- Tabakalı Örneklem: Öncelikle kitle belirli bir çerçevede gruplandırılır ve her bir gruptan rasgele örneklem çekilerek birleştirilmesiyle elde edilen yöntemdir. Burada önemli olan gruplar kendi içinde homojen, gruplar arası ise heterojen olmalıdır. Örneğin, PAÜ deki öğrencilerin aylık ortalama harcamaları için kitleyi gruplandırılabilir. Tıp, Mühendislik, Fen Edebiyat, İktisadi ve İdari Bilimler, Mimarlık ve Tasarım gibi fakültelerden yoğunluklarına göre her birinden rasgele öğrenciler seçip örneklem oluşturmak.
- Küme örneklem: Kitleyi ayrı ayrı temsil edebilen kendi içerisinde heterojen, aralarında homojen yapıya sahip kümelerin seçimine bağlı yöntemdir. Örneğin, sanayi bölgesinde tekstil fabrikalarında çalışan işçilerin çalışma koşulları üzerine araştırma yapılmak istensin. Aynı koşullara sahip her bir fabrika birer küme olmuş olup rasgele fabrikalar seçilerek örnekleme oluşturma yöntemidir.

1.3. Değişken ve değişkenlerin ölçülme düzeyleri

Değişken: Birimlerin farklı değerler alabildikleri nitelik veya niceliklerdir. Örneğin müşterilerin oluşturduğu bir kitlede; meslek, eğitim düzeyi, cinsiyet, yaş, aylık gelir gibi.

Nitel (Kategorik) değişkenler: Gözlemlerin sınıflara ayrılarak açıklandığı ölçümlerdir. Cinsiyet, eğitim durumu, kan grubu gibi.

Nicel (Sayısal) değişkenler: Sayısal değerlerden oluşan ölçümlerdir. Boy, kilo, yaş, nüfus gibi. Sayısal değişkenler *kesikli* ve *sürekli* olmak üzere ikiye ayrılır. *Kesikli değişkenler*, belirli aralıkta tam değerler alabilen değişkenlerdir. Sınıftaki öğrenci sayısı, otoparktaki araba sayısı gibi. Sürekli değişkenler ise belirli aralıkta tüm değerleri alır. Ağırlık, uzunluk gibi.

Değişkenlerin ölçümleri genel olarak dört farklı şekilde yapılabilir.

- **Sınıflama Ölçeği:** Gözlemlerin sadece sınıflanması veya etiketlenmesiyle sınırlıdır. Sınıflar arasında büyüklük ya da sıra açısından bir ilişki bulunmaz. Kategorik

değişkenlerdir. Örneğin, Cinsiyet, kan grubu, meslek grubu, Gözlük kullanımı (0: Gözlüksüz, 1: Gözlüklü) gibi.

- **Sıralama Ölçeği:** Gözlemlerin belirli bir kritere göre sıralayarak etiketlenmesidir. Değişken değerleri bakımından önem sırası bulunur. Kategorik değişkenlerdir. Örneğin, Eğitim durumu (ilkokul, ilköğretim, lise, üniversite), Üniversite öğretim üyesi unvanları ve askeri rütbelere, ekonomik durum (1: çok kötü, 2: kötü, 3: orta, 4: İyi, 5: çok iyi).
- **Eşit aralıklı Ölçek:** Bu ölçekte üzerinde durulan değişken belirli iki değer arasında sonsuz değer alabilir. Bu ölçekteki 0 değeri, ölçülen karakteristiğin olmadığını göstermez. Aynı şekilde ölçüm karakteristiklerinden biri diğerinin katlarıyla ifade edilemez. Sayısal değişkenlerdir. Örneğin, Sıcaklık ölçümleri. 0 derece, sıcaklığın olmadığını göstermez ve 4 derece, 2 derecenin iki katı değildir.
- **Oran Ölçeği:** Zayıftan kuvvetliye doğru sıraladığımızda yukarıdaki ölçeklerin en hassas olanıdır. Ölçülen karakteristiğin 0 olması o karakteristiğin olmadığını gösterir. Aynı şekilde ölçülen bir karakteristik diğerinin katları ile ifade edilebilir. Sayısal değişkenlerdir. Örneğin; ağırlık, uzunluk, hız, gelir gibi. Sıfır ağırlığa sahip olmak yokluğu ifade ettiği gibi 50 kg'ın iki katı 100 kg ifade eder. Tüm matematiksel işlemler yapılabilir.

2. FREKANS TABLOLARI

Her bir gözlemin veri setinde kaç kez tekrarlandığı ya da belirli bir aralıkta gözlemlerden kaçının bulunduğunu gösteren tablolara *frekans tabloları* denir. f ile gösterilir. n gözleme sahip veri için j inci sınıfın frekansı f_j ile, göreceli frekansı da f_j/n ile gösterilir. Birikimli frekans ise kesikli değişken değerlerinin x_i 'ye eşit ya da daha küçük değerlerine karşılık gelen frekansları toplamıdır. Yani;

$$(\text{Birikimli frekans}) F_i = \sum_{j=1}^i f_j.$$

Bir veriyi özetlemek ve daha anlaşılır hale getirmek istediğimizde frekans tabloları kullanırız. Veride her bir X değerinin karşısına o değerin frekansı yazılır. Mesela 150 birimlik veride 5 tane farklı birim varsa bu 150 birim 5 sınıf halinde özetlenmiş olur. Farklı birimlerin çok fazla olması halinde her bir sınıf için belirli aralıklar oluşturularak bu aralığa düşen gözlem sayıları frekans olarak yazılır.

Örnek: Bir sitedeki 15 ailenin çocuk sayıları aşağıda verilmiştir. Frekans tablosu oluşturunuz.

0	2	1	3	1	2	1	2	2	1	3	0	4	3	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Çocuk sayısı	Frekans	Birikimli frekans	Görel frekans
0	2	2	$\frac{2}{15} = 0.133$
1	5	7	$\frac{5}{15} = 0.333$
2	4	11	$\frac{4}{15} = 0.267$
3	3	14	$\frac{3}{15} = 0.200$
4	1	15	$\frac{1}{15} = 0.067$

0 grupta kaç eleman var

Kendi'si + öcesi

frekansı

toplam

➤ Sınıflandırılmış Frekans tablosu: İlgilenilen veri setinde sürekli değişken ya da fazla sayıda kesikli değişken içeren gözlemler varsa, değişkenin değer aralığı alt aralıklara bölünüp sınıflar oluşturulur ve her bir sınıfın frekansı hesaplanarak tablo oluşturulur. Sınıflandırılmış frekans tablosu için aşağıdaki adımlar sırası ile uygulanır.

- Veri seti için n =gözlem sayısı, EB=en büyük değer, EK=en küçük değer belirlenir.
- Sınıf sayısı k keyfi ya da $k \geq \sqrt{n}$ olacak şekilde seçilir.
- Sınıf genişliği $h \geq \frac{EB-EK}{k}$ ile hesaplanır.
- En küçük gözlem değeri ilk sınıfın alt limiti olarak seçilir ve bu değere sınıf genişliği ardışık olarak eklenerek diğer sınıfların alt limitleri oluşturulur.
- Sınıflara düşen gözlem sayıları yani frekanslar belirlenir.

Örnek: 20 kişiye yapılan ankette kişilerin yaşları sorulmuş ve aşağıdaki yanıtlar elde edilmiştir. Sınıflandırılmış frekans tablosunu oluşturunuz.

35	35	32	40	48	28	35	32	35	52
48	41	47	35	22	33	34	52	20	22

- $n=20$ ve $EB=52$ ve $EK=20$
- Sınıf sayısı $k \geq \sqrt{20} = 4.5$ olup $k = 5$ alınır.
- Sınıf genişliği $h \geq \frac{52-20}{5} = 6.4$ olup $h = 7$ alınır.
- Sınıf alt limitler: 20, 27, 34, 41, 48 dir.
- Sınıflara düşen frekanslar tabloda verilir.

Yaş (X) için sınıflar	Frekans	Birikimli frekans	Görel frekans
$20 \leq X < 27$ (20-27)	3	3	$\frac{3}{20} = 0.15$
$27 \leq X < 34$ (27-34)	4	7	$\frac{4}{20} = 0.20$
$34 \leq X < 41$ (34-41)	7	14	$\frac{7}{20} = 0.35$
$41 \leq X < 48$ (41-48)	2	16	$\frac{2}{20} = 0.10$
$48 \leq X < 55$ (48-55)	4	20	$\frac{4}{20} = 0.20$

Örnek: Bir sitedeki 27 konutun aylık doğalgaz tüketimi verileri aşağıdaki gibidir. Sınıflandırılmış frekans tablosunu oluşturunuz.

300	390	365	420	432	490	595	540	460	390	420	430	490	480
490	390	405	425	450	420	440	420	500	460	570	480	540	

- $n=27$ ve $EB=595$ ve $EK=300$
- Sınıf sayısı $k \geq \sqrt{27} = 5.2$ olup $k = 6$ alınır.
- Sınıf genişliği $h \geq \frac{595-300}{6} = 49.2$ olup $h = 50$ alınır.
- Sınıf alt limitler: 300, 350, 400, 450, 500, 550 dir.
- Sınıflara düşün frekanslar tabloda verilir.

Doğalgaz tüketimi	Frekans	Birikimli Frekans
300-350	1	1
350-400	4	5
400-450	9	14
450-500	7	21
500-550	3	24
550-600	3	27

