VERİLEN ANALİZİ VE YORUM

(En Çok Kullanılan Bazı Tekniklerle Birlikte)

Doç. Dr. Şefik UYSAL

Bütün araştırmalar yüzeyde görülebilen olguların arkasında yatan ve kolaylıkla görülemeyen ilişkiler sistemini ortaya çıkarmak amaç ve isteği ile ele alınmaktadır. Böylece mevcut bilgilere yenileri eklenmektedir. Dikkatlice hazırlanmış bir araştırma projesinin, en çok yaratıcılık ve yeterli alan bilgisi isteyen aşaması şüphesiz, araştırma için toplanan verilerin dikkatli ve sistematik bir biçimde analizi ve bulguların yorumlanması aşamasıdır. Bu aşama, araştırma yapanların yaşantılarında görüldüğü gibi, güçlüklerle dolu olduğu kadar, mevcut ve bilinmeyen yeni bulgulara doğru yönelindiği zaman aynı oranda da önemli bir motivasyon kaynağı olmaktadır.

Araştırma, hiç bir zaman bir takım bilimsel yöntemler uygulayarak veri toplamak biçiminde anlaşmamalıdır. Önemli olan toplanan ve Sosyal Bilimlerde özellikle olduğu gibi, elde mevcut geniş bir veri topluluğu içinde araştırma amaçları yönünde anlamlı görülen ilişkileri yakalayabilmek için gösterilen yaratıcı uyanıklıktır. Doğal olarak, analiz aşamasını, araştırmanın diğer aşamalarından ayrı düşünmek mümkün değildir. Böyle olduğu takdirde bulgular ve yorumların gerçeği yansıtmasına şüphe ile bakılabilecektir.

Bilindiği gibi araştırmalar çoğu kere geniş bir bilgi sistemi içinde bilinmeyen bir takım sınırlı ve belirli sorunlara yöneliktir. Dolayısı ile bulgular o alanda mevcut bilgilerin içinde yerine konulduğu zaman yeni araştırmalara ihtiyaç duyulması doğaldır. Analiz bu yüzden de, yeni araştırmalara doğru bir ilk adım olarak düşünülmektedir.

İnsan zekâsının üretebileceği bir çok sorulara cevap bulmak ve araştırıcıların ilgi duydukları süreçleri açıklamak maksadı ile çeşitli analiz yöntemleri düşünülebilir. Bu yöntemleri tek tek ele alarak incelemek yerine, hangi yöntem olursa olsun söz konusu edilecek temel kavramları tanımak daha yararlı olacaktır. Her tür analiz ve yorumda şu dört husus açıklığa kavuşturulmalıdır. 1) Araştırmada konu edinilen değişkenler, 2) Toplanan verilerin türü, 3) Nicel verilerin toplanmasında kullanılan ölçekler, ve 4) Toplanan verilerin analizinde yardımcı olan istatistik teknikler. Bu tartışma, yukarda belirtilen dört sorunu ele almak sureti ile analiz ve yorum aşamasında gerekli temel bilgilere değinmiş olacaktır.

1). Değişkenler

Değişken şöyle tanımlanmaktadır: Bireyden bireye veya objeden objeye değişik değerler alan özellik ya da durumlara değişken denmektedir. Bu değişiklik iki şekilde düşünülebilir. Birincisi, bireyin ya da objenin belirlenmiş bir özelliğini iki ayrı zamanda gözlemek suretiyle ile olur. Bireylerin politik davranışlarının, milletvekili seçiminden önce ve sonra gözlenmesi bu türe en uygun bir örnek olabilir. Burada değişken "politik davranış"tır ve iki ayrı zamana göre bir değişiklik göstermesi beklenmektedir.

İkincisinde ise, birden fazla birey ya da objelerin, belirli özellikleri açısından, aynı zamanda gözlenmesi şeklinde olmaktadır. Örneğin, ilkokula yeni kaydolan birinci sınıf öğrencilerinin ağırlıklarının ölçülmesi ya da Türkiye'deki illerde belli bir ayda alınan yağış miktarı gibi. İlk örnekteki değişken "ağırlık"tır. Burada öğrenciler arasında bir farklılık görülecektir. İkinci örnekte ise "yağış" bir değişken olup, bu da bir ilden diğerine miktar açısından bir farklılık verecektir.

Yukardaki açıklamalardan da anlaşılacağı gibi değişkenleri tek bir kategoride düşünmek, sorunu çok genel açıdan almakla mümkündür. Özellikle bilgilerin analizi bakımından değişkenleri bazı türleri ile beraberce tanımakta fayda vardır.

Değişkenler önce sürekli ve süreksiz, sonra da bağımlı ve bağımsız değişkenler olmak üzere iki genel kategoride tartışma konusu yapılacaktır.

a- Sürekli (continuous) ve Süreksiz (discrete) Değişkenler

Değişkene ilişkin ölçmede kullanılan üniteleri kavramsal olarak sonsuza giden alt ünitelere ayırabilmek söz konusu olduğu takdirde buna sürekli değişken denir. Örneğin boy bir sürekli değişkendir. Çünkü bireyin boyu gelişme süreci içinde bir yaştan diğerine belli üniteleri tamamlayarak geçer ve ölçmede bu anlamlı olarak gösterilebilir Bunu sadece ölçü aracının hassasiyeti sınırlayabilir. Burada ölçme üniteleri metre, desimetre, santimetre, milimetre ve mikrometre olarak düşünüldüğü zaman bireyin boyundaki değişmeler, bir en küçük üniteden diğerine geçmek sureti ile bir süreklilik gösterir.

Diğer taraftan bazı ölçmelerde daha küçük ünitelere gitmek ve ayrıntılarını ortaya koymak mümkün değildir. Bu tür değişkenlere de süreksiz değişkenler adı verilmektedir. Örneğin sınıf mevcudu, ailedeki çocuk sayısı, cinsiyet veya medenî durum süreksiz değişkenlerdir. Görüldüğü gibi burada, obje veya bireyler sözü edilen özelliklere göre birbirlerinden farklı kategorilerdedir. Birey, medenî durumuna göre evli ise, aynı birey bekâr kategorisine giremez. Aynı durum sınıf öğrenci sayısı için de düşünülür. Öğrenci mevcut 40 olan bir sınıf için öğrenci mevcudu 40 3 /4 şeklinde düşünülemez. Diğer bir deyişle, bireyler ya da objeler ele alınan özelliklerde birbirlerinden ayrılmış kategorilerde görünmektedir.

Bu tartışmada hatırlanması gereken önemli bir husus şudur: Sürekli ve süreksiz değişken ayırımı ölçmeden değil, ölçülen özellikten yani değişken kavramından ortaya çıkmaktadır. Bir zekâ ölçeği uygulandığını düşünelim ve birey X'in puvanı 120 iken birey Y'nin 122 olduğunu kabul edelim. Eğer puvan açısından alınırsa, burada "zekâ" bir süreksiz değişken olarak düşünülebilir. Çünkü ölçme sonucu, puvanlar tam sayılar halinde belirtilmektedir. Halbuki, farklılık ölçülen özellik açısından ele alındığı takdirde zekâ bir devamlı değişkendir. O halde ölçme sonuçları, aslında, devamlı değişkenleri de belirli ölçülerde sınırlamış olmaktadır. Örneğin hiç bir zekâ ölçeği sonuçları, iki puvan arasını en küçük ünitelere ayırmak sureti ile ayrıntılı puvanlar halinde vermez.

Sürekli ve süreksiz değişkenlerin araştırma sorunları açısından tartışılmasında bir diğer önemli noktaya da değinmek gerekir. Sürekli değişkenlerin ölçülmesinde tam bir isabet sağlamak hiç bir zaman mümkün değildir. Çünkü, ölçü araçlarının güvenirlik ve geçerlik sorunları ile ilgili bölümde de ayrıntıları ile tartışıldığı gibi, sürekli değişkenlerin ölçülmesinde her zaman bir hata söz konusudur. Halbuki, aynı sorun süreksiz değişkenler için mevcut değildir. Çünkü, örneğin, bireyin evli ya da bekâr oluşunu saptamak her durumda mümkündür. Bir ailenin çocuk sayısını hatasız bilmek hiçte güç olmamalıdır.

Sürekli değişkenlerin ölçümesindeki bu güçlük yüzünden araştırıcılar genellikle verileri, saptanmış bazı özelliklere göre gruplama yöntemi uygulamaktadırlar. Böylece hatanın yoğunlaşmasını küçümsenmeyecek oranda kontrol altına almış olmaktadırlar. Örneğin, zekâ ölçeği sonuçlarını bir dağılımda gösterirken puvanları 80–84, 85–89 v.b. şeklinde gruplamaları gibi.

b. Bağımsız (independent) ve Bağımlı (dependent) Değişkenler

Bu tür değişkenler daha ziyade deneysel araştırmalar için söz konusudur. Deneylerde, araştırıcı bazı özellik ya da koşulları sabit tutar ya da kontrol altına alır. İşte bu özellik veya koşullar bağımsız değişken olarak nitelendirilir. Bazı literatürde buna kontrol değişkeni adı da verilir.

Araştırıcının deney sonunda değişeceğini beklediği koşullar ya da davranışlara ilişkin değişkenler de bağımlı değişkenler olarak adlandırılır.

Bir örnekte bu değişkenleri görelim. "Öğrencilerin zekâ seviyeyelerinin akademik başarılarında önemli bir etmen olduğu" hipotezine dayanan bir deneysel araştırmada, "zekâ seviyesi" bağımsız ve "akademik başarı" ise bağımlı değişken olmaktadır. Çünkü öğrencilerin akademik başarıları, onların zekâ seviyelerine göre değişen bir değişken olmaktadır.

Bağımlı ve bağımsız değişikenler her araştırmanın benimsediği hipotez ya da sorunlara göre değişebilir. Yukarda verilen örnekte "zekâ seviyesi" bağımsız değişken olmuştur. Bu durum "zekâ seviyesi" her araştırmada bağımsız değişken olacaktır anlamında düşünülmemelidir. "Çocukların 8 yaşına kadar içinde bulundukları toplumsal koşullar, onların zekâ seviyelerinin oluşumunda önemli bir etmendir" hipotez'ni doğrulamak için yapılan bir araştırmada ise "zekâ seviyesi" bağımlı, "toplumsal koşullar" ise bağımsız değişken olarak görünmektedir.

2). Toplanan Verilerin Türleri

Genel hatları ile değişkenleri tanıdıktan sonra, değişkenler hakkında analizde kullanılacak bilgileri de sınıflamak yararlı olacaktır. Çünkü bütün analizlerde, değişkenlerin saptanmasından sonra ilk önce ele alınan konu toplanan verilerin türüdür. Araştırmada toplanan bilgileri iki genel kategoride düşünmek mümkündür; 1) nitel, 2) nicel, Şimdi bu iki veri türünü genel hatları ile tanıyalım.

Araştırma konusu olan gözlemde, bireyler ya da objeler sayısal bir ölçekte belirtilmişse nicel, belirtilmemişse nitel olmaktadır. Diğer bir deyişle, bireylerin ya da objelerin ölçülen özellikleri sayısal kavramları içeriyorsa bu, nicel bir ölçekten elde edilen bir veri topluluğunu vermektedir. Aksi ise ise nitel bir veri topluluğu demektir.

Bir örnek verelim.

Üniversite öğretimi ile ilgili çeşitli sorunları saptamayı içeren bir araştırmada, bireylerin tavırlarının şu tür ifadelerle ölçülmekte olduğunu düşünelim.

	Katılırım		Kararsızım		Katılmam
	1	2	3	4	5
Bugünkü öğretim toplumun istihdam politikasına uygundur.	( )	( )	( )	( )	( )

Birey kendi sahip olduğu görüşünü böyle bir ölçekte belirtecektir. Burada dikkat edilmesi gereken husus sayıların kavram olarak neyi belirttikleridir. Sayılar bir sembol olarak düşünülmemelidir. Nitekim bu ölçekte "katılırım" görüşü "1" olarak gösterildiği gibi "5" te olabilir ve diğerleri buna göre düzenlenebilir. Görüldüğü gibi örneğin "2" "katılırım" görüşüne yakın bir kararsızlığı, "4" ise "katılmam"a yakın bir kararsızlığı belirtmektedir.

Aynı ifade;

Bugünkü öğretim toplumun istihdam politikasına uygundur.

( ) 1. Evet, doğrudur

( ) 2. Hayır, doğru değildir.

biçiminde de düzenlenebilir. Bu kere nitel bir ölçü ortaya konmaktadır. Dikkat edileceği gibi, buradaki yazılı "1" ve "2" sadece davranışı belirten, tepkiyi belli sembollere bağlayan sayılardır.

Görüldüğü gibi, nicel ve nitel farklılığı, tepkilerin sayılarla gösterilmesi veya gösterilmemesi biçiminde ele almamak gerekir. Eğer sayılar yalnızca tepkilerin kategorileşmesinde kullanılıyor ise bu nitel, fakat eğer tepkileri kategorileştirmenin yanında, bu tepkilerin kendi aralarında ilişkilerini dikkate alan bir sıralama da yapılıyor ise bu defa nicel bir bilgi topluluğu söz konusudur.

Genel hatları ile veri türlerini tanıdıktan sonra, son yıllarda Sosyal Bilim araştırmalarında önem kazanan nicel verilerin hangi ölçeklerle elde edilebileceğine değinelim. Diğer bir deyişle nicel veri türlerini açıklayalım.

3). Nicel Veri Türleri

Sosyal Bilim araştırmalarında analize hazır olan veriler çeşitli ölçekler yolu ile toplanmaktadır.

a. Sınıflama Ölçekleri (Nominal Scale)

Sosyal Bilimlerde veri toplanırken kullanılan en basit ölçekler nominal ölçeklerdir. Bunda, yapılan işlem bireyleri, objeleri veya durumları kategorilere veya sınıflara ayırmaktır. Örneğin cinsiyet değişkeni ile ilgili bir anket sorusunu (1) erkek (2) kadın diye gruplayıp numaralamak bu tür bir ölçektir. Böyle bir ölçmedeki esas düşünce eşit (=) veya eşit değil (≠) şeklinde bir ayırıma gitmektir. Görüldüğü üzere bu ölçeklerde yapılan iş kategoriler ya da sınıfların birbirlerinden farklı olduklarını belirlemekten ibarettir. Dikkat edileceği gibi, kateğori ya da sınıfları belirten bu sayılarla herhangi bir matematiksel işlem yapmak mümkün değildir. Sadece her kategori ya da sınıftaki sayılardan (frekans) yararlanarak mod saptanabilir.

b. Sıralama Ölçeği (Ordinal Scale)

Belirli özelliklere göre obje veya bireyleri sıraya koymada kullanılan ölçekler sıralama ökçekleri adını alır. Böylece daha büyüktür (>) veya daha küçüktür (<) demek mümkün olabilmektedir.

Örneğin bir sınıfta uygulanan başarı testinde, öğrenciler aldıkları puvanlara göre en büyükten en küçüğe, veya aksi bir sıraya konulurlar ve her sıraya bir sayı verilir. Anlaşılacağı gibi böyle bir sıralamada, ölçülen değişken açısından, bireyler arasındaki farkın derecesini görmek ve her bireyin sahip olduğu miktarı gruba göre bilmek mümkün olmamaktadır. Bu yüzden bu tür bir veri için ancak, uzaklıkları ya da sıralar arası farkları karşılaştırarak, farkın miktarına değinmeksizin "daha büyük", "daha küçük" gibi ifadeler kullanılabilir.

Sıralama ölçekleri kullanmak sureti ile elde edilen veriler için herhangi bir matematiksel ve istatiksel işlem için çok dikkatli olmak gerekir. Belki, ancak bir miktar hatayı göze alarak, ortalama ve standart kayma işlemleri uygulanabilir.

c. Aralıklı Ölçekler (İnterval Scale)

Eğer, ölçek gerçek bir sıfır noktası, yani bir başlangıç noktasına göre bireylerin her birinin uzaklığına ait yorum olanağı vermiyor, fakat bireyler veya objeler arasındaki farklılıkların miktarı ya da uzaklığı yönünde bir yorum olanağı kazandırıyor ise buna da aralıklı ölçekler adı verilmektedir. Bu tanımda iki husus dikkati çekmelidir. Bunlardan birincisi, başlangıç noktası sorunudur. Aralıklı ölçeklerde gerçek bir sıfır noktası olmamasına karşın, bireyler arası farklılıkları söz konusu edebilmek için bağıl bir başlangıç noktası esası kabul edilmektedir. Her araştırıcı, araştırma amacına göre bağıl bir başlangıç noktası seçebilmektedir.

İkinci husus ise, ölçekte seçilen başlangıç noktasının altında ve üstünde kalan kısımlarının eşit ünitelere bölünmüş olmasıdır. Aralıklı ölçeklerde de bu husus belli yöntemlerle sağlanır.

Aralıklı ölçeklerde kullanılan bağıl sıfır noktasına göre yapılan yorumlara özellikle dikkat etmek gerekir. Bir örnekte ortaya çıkan durumu görelim. Diyelim ki bir zekâ ölçeği uygulandı ve bu ölçek için bağıl başlangıç noktası (ortalama) 50 olarak alındı. Birey A'nın bu başlangıç noktasına göre puvanı 25 ve Birey B'nin ise 50 olsun. Bir an için eşit üniteler kullanabildiğimizi de düşünürsek birey B'nin puvanının birey A'nın iki katı olduğu söylenebilir. Ama eğer başlangıç noktası olarak kabul ettiğimiz 50 puvanı da dikkate alırsak A'nın puvanı 50+25=75, ve B'nin ise 50+50=100 olmaktadır. Görülüyor ki ilk yorum bu sonuç için geçerli kabul edilemez.

Mamafih, aralıklı ölçekler için bir çok istatistik işlemler kullanmak mümkündür. Örneğin ortalama, standart kayma ve Pearson Korelasyonu gibi.

d. Oranlı Ölçekler (Ratio Scale)

Görülüyorki gerçek bir başlangıç noktası, diğer deyimle sıfır başlangıç noktası ve eşit ünitelerle bir ölçme yapabilmek ihtiyacı çoğu kere hissedilmektedir. İşte bu tür bir ölçek gerçekleştirilebildiği takdirde buna oranlı ölçekler denmektedir. Bu ölçekler, objeler ya da bireyler arası eşitlikleri belirtebilmek, onları sıraya koymak, aralarındaki farkın miktarını belirtmek ve bu farklılıkları oranlar halinde ortaya koymak bakımından önemli yararlar sağlamaktadır. Bu tür bir ölçek kullanmak sureti ile elde edilen bilgilere bütün matematik ve istatistik işlemler rahatlıkla uygulanabilir.

Bilindiği gibi fiziksel ölçümlerde her zaman gerçekleştirilebilen bu tür ölçeklerin Sosyal Bilimlerde söz konusu olan davranışsal özelliklerin ölçümünde nadiren gerçekleştirilebilmektedir. En iyi standardize edilmiş bir ölçekte dahi bağıl başlangıç noktası uygulanabilmektedir. Bu yüzden Sosyal Bilim araştırmalarında elde edilen bilgiler çoğu kere aralıklı ve bazen de sıralama ve sınıflama ölçekleri ile elde edilmektedir.

Buraya kadar, analiz için toplanan nicel verilerin hangi tür ölçekler yolu ile elde edilebileceği yönünde genel bir bilgi verildi. Görülüyor ki, analiz aşamasına gelmiş olan bir araştırıcı, hangi ölçme sonucu elde edildiğine dikkat etmek sureti ile hem kullanabileceği istatistik teknikleri saptamada ve hem de yapabileceği yorumun sınırları hakkında isabetli kararlar verebilecektir.

4). Araştırmada Kullanılan Bazı İstatistik Kavramlar

Sosyal Bilim araştırmalarında araştırıcıyı en çok sevindiren çeşitli araçlar yolu ile toplanmış çok geniş bir veri topluluğundan, araştırma amaçları yönünde bazı bilimsel bulgulara varabilmektir. Fakat diğer taraftan da bilinirki toplanan bilgileri oluşturan olgular ve sayılar tek başlarına alındıkları zaman anlamsız olmaktadırlar. Bu olgular ve sayılar ayni veri topluluğu içindeki diğerleri ile karmaşık bir ilişki düzeni içindedirler. Olgular göründüğü kadar kolay anlaşılmazlar. Çünkü çoğu kere her olgu subjektif ya da objektif, değişen derecelerde ve kombinasyonlarda, bir çok elementleri içermektedir. İşte bu yüzden araştırmada toplanan verilerin, analizine imkân hazırlayabilmek için sistematik bir düzenlemeye gidilmesi bir ön koşuldur. Sosyal Bilimlerde, özellikle son yıllarda yapılan araştırmalarda toplanan veriler çoğu kere nicel özellikler göstermektedir. Bu böyle olduğuna göre toplanan çok geniş verilerin analiz ve yorumunu kolaylaştırmak maksadı ile istatistik tekniklerden geniş ölçüde yararlanmak gereklidir. Ayrıca istatistik, analize bilimsel bir nitelik te kazandırmaktadır. Bu tartışmamızda bazı temel ve oldukça yaygın kullanılabilen istatistik teknikleri hatırlatmak maksadı ile fazla ayrıntılara inmeden yer vermek yararlı olacaktır.

a. Verileri Tablolama

Verileri tablolar halinde gösterebilmenin ilk koşulu bu verilerin sistematik bir sınıflandırılmasının yapılmasıdır. Sosyal Bilim araştırmalarında toplanan verilerin sınıflandırılmasında aşağıdaki kıstasların bir veya birkaçı, oldukça yaygın kullanılagelmektedir.

i. Coğrafi bölgeler: Verileri coğrafî bölgeler, kentsel ya da kırsal bölgeler v.b. gibi kıstaslara göre sınıflamak

ii. Zaman: gün, hafta, ay ve yıl gibi zaman birimlerine göre sısıflamak

iii. Nitelik: bilgi toplanan bireylerin niteliksel özelliklerine göre sınıflamak, örneğin, cinsiyet, meslek, evlilik durumu v.b. gibi.

iv. Nicelik: burada düşünülen niceliksel verilerin dikkate alınarak sınıflandırılmasıdır. Örneğin, bireylerin bir ölçü aracında aldığı puvanlara ve benzeri istatistiksel bilgiler gibi.

Verilerin tablolanması önce tek bir değişkene göre düzenlenir ve buna yan dağılım (marginal distribution) denir. Bundan sonra birbirleri ile ilişkileri görülmek istenen iki ve en çok dört değişkene göre çapraz tablolar (cross-tables) düzenlenir. Bu tablolar çoğu kere frekanslar yanında ya sûtun veya satır toplamlarına göre yüzdeleri de içermektedir. Elde edilen dağılımlar frekansı da içerdiği için aynı zamanda frakans dağılımı olarakta nitelendirilmektedir.

b. Grafikler

Analizde, frekans dağılımlarını daha anlaşılır hale getirmek için bunları grafikle göstermek en yaygın kullanılan bir yoldur. Burada bunlardan en çok kullanılan üç tanesi söz konusu edilecektir.

i. Histogram (bar grafiği)

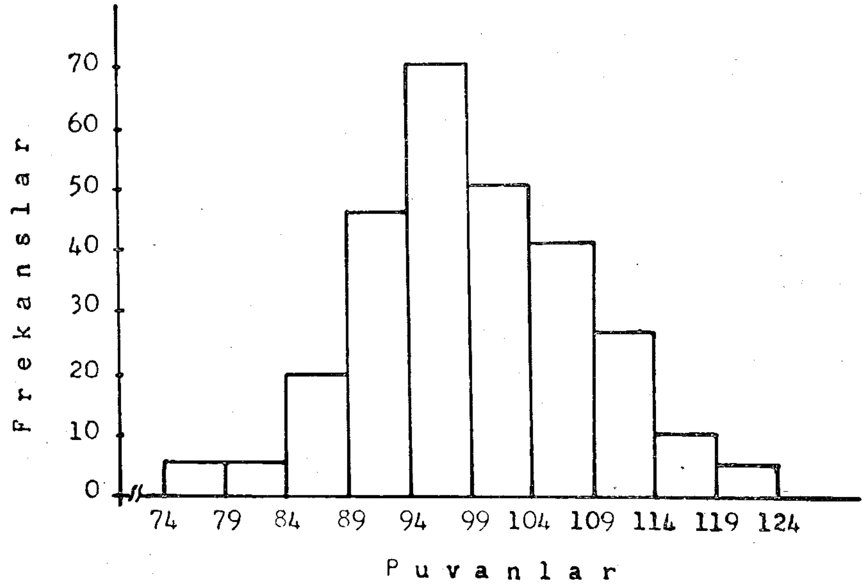
Histogram frekansların sütünlar halinde gösterilmesidir. Bunda freakanslar alanlar halinde belirtilmektedir. Bir örnekle durumu açıklayalım. Bir okulun 300 öğrencisine bir zekâ ölçeği uyguladığımızı düşünelim. Öğrencilerin aldığı puvanlara göre sayıları şöyle bir frekans dağılımı vermektedir.

TABLO 1.

Frekans Dağılımı (Gruplanmış seri)

Puvan aralıkları	Frekans	Yığmalı fre.
120 - 124	5	270
115 - 119	10	265
110 - 114	25	255
105 - 109	40	230
100 - 104	50	190
95 - 99	70	140
90 - 94	45	70
85 - 89	15	25
80 - 84	5	10
75 - 79	5	5
Toplam	270	

Görüldüğü gibi histogramda puvanlar eşit aralıklar halinde belirtilmekte ve her sütûn o aralığa isabet eden freakans toplamını vermektedir.



Şekil 1. Bar Grafiği

ii. Frekans Poligonu

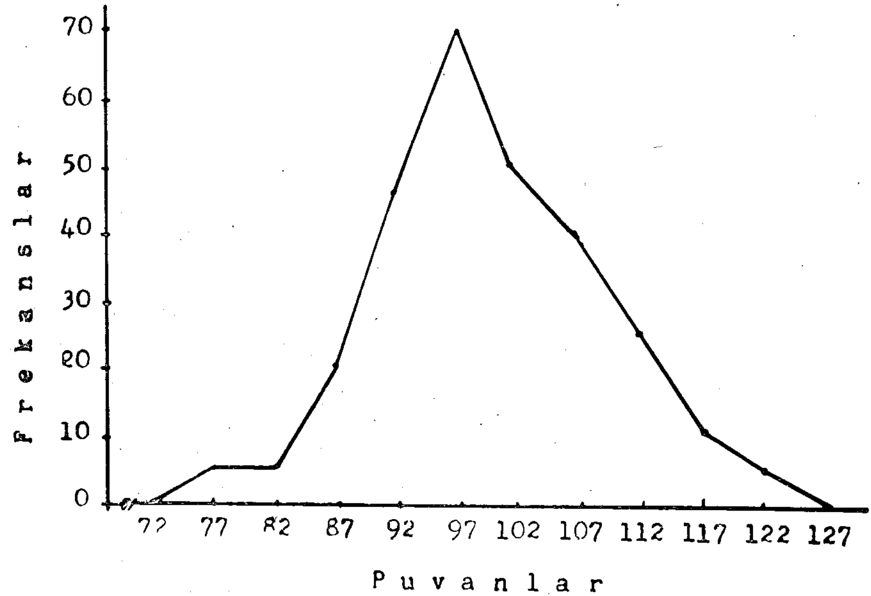
Frekans poligonu bir aralığa düşen frekansın o aralığın orta noktasında toplandığı varsayımına göre, bar grafiğindeki sütunların tepe orta noktalarının bir çizgi ile diğerlerine birleştirilmesi sureti ile elde edilir.

Tablo 1'deki bilgilerden yararlanarak çizilen frekans poligonu Şekil 2'de gösterilmiştir.

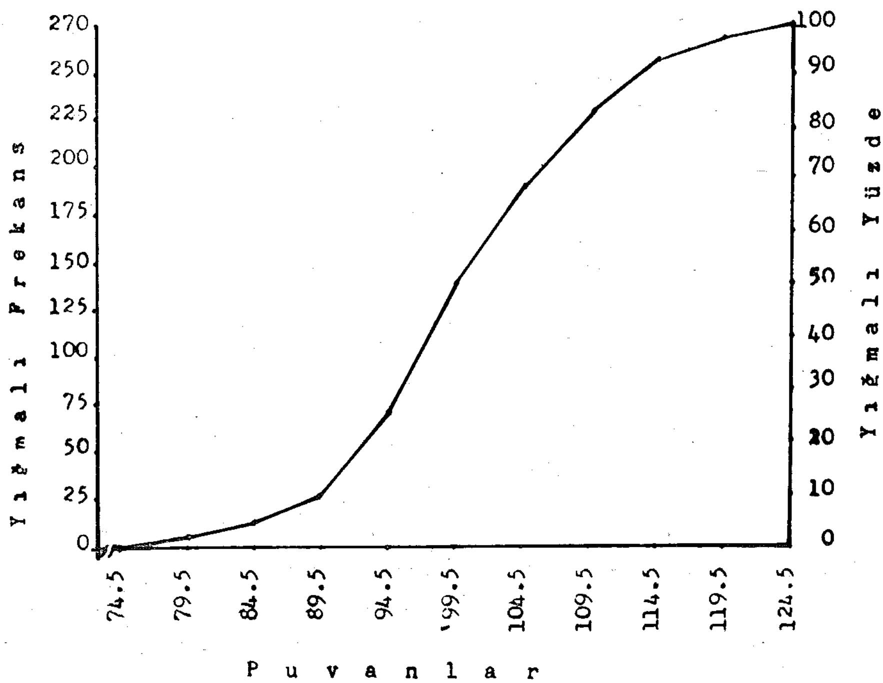
iii. Yığmalı Grafiği

Bu grafiğin çizilmesinde, grafiğin frekansları gösteren ekseninde bu defa yığmalı freakans gösterilir. Yığmalı frekansı elde etmek için en alt aralıktaki freakanstan başlanarak bir üst aralığın frekansları en üst aralığa kadar eklenerek elde edilir. Grafiğin X ekseninde ise her aralığın üst gerçek sınırları belirtilir. Yine Tablo 1'deki bilgilerden yararlanarak çizilen yığmalı grafik Şekil 3'de gösterilmiştir. Yığmalı grafik, aynı zamanda "S-Grafiği" ya da "Ogive" olarakta adlandırılır.

Yığmalı grafiği, yüzdenler içinde kullanmak mümkündür. Bunun için de Y ekseninin karşısına çizilen kenarda yığmalı yüzdeleri vererek, hem yığmalı frekansı ve hem de yığmalı yüzdeyi ayni grafik üzerinde okumak tercih edilmektedir.



Şekil 2. Frekans Poligonu



Şekil 3. Yığmalı Grafiği

Bu grafiklerin dışında çeşitli değerlerle gösterilen bir çok grafik türleri düşünmek te mümkündür. Unutulmaması gereken bir husus, grafiklerin toplanmış olan verilerin anlaşılmasına kolaylık getirdiği gibi, bir çok ifadelerle açıklanabilecek hususların öz olarak aktarılmasını sağladıklarıdır. Bu yüzden hangi tür grafik kullanılırsa kullanılsın anlaşılması kolay, öz ve aktarılması gereken bilginin mahiyetine uygun olmalıdır.

c. Merkezi Eğilim Ölçüleri

Sosyal Bilim araştırmalarında toplanan çok geniş verileri özetlemek ve tek bir sayısal değere indirgemek bakımından merkezî eğilim ölçüleri (vasat ölçüleri) özel bir önem taşımaktadır. Çeşitli merkezî eğilim ölçüleri vardır. Fakat biz bunlardan en çok kullanılan mod, medyan ve aritmetik ortalamayı kısaca özetleyeceğiz.

i. Mod

Bir dağılımda en çok tekrar eden gözlem değeri (puvan) mod'dur. 10 kişiye bir ölçek uyguladığımızı ve puvanların şöyle olduğunu düşünelim. 47, 49, 50, 51, 54, 54, 54, 57, 57, 60. Burada en çok tekrar eden 54'tür. Çünkü 3 kişi bu puvanı almıştır.

Eğer puvanlar ya da gözlem değerleri gruplanmışsa işlem biraz farklı olmaktadır.

Tablo 1'deki dağılımın mod ve diğer merkezi eğilim ölçüleri için gerekli bilgilerle Tablo: 2'de verelim.

TABLO 2.
Merkezi Eğilim ölçüleri için gerekli
Bilgilerle Frekans Dağılımı

Aralıklar	Frekans	Yığmalı Frekans	Orta Nokta	Alt ve üst gerçek sınırlar
120 - 124	5	270	122	119.5–124.5
115 - 119	10	265	117	114.5 - 119.5
110 - 114	25	255	112	109.5 - 114.5
105 - 109	40	230	107	104.5 - 109.5
100 - 104	50	190	102	99.5 - 104.5
95 - 99	70	140	97	94.5 - 99.5
90 - 94	45	70	92	89.5 - 94.5
85 - 89	15	25	87	84.5 - 89.5
80 - 84	5	10	82	79.5 - 84.5
75 - 79	5	5	77	75.5 - 79.5

Bu dağılımda mod frekansı en yüksek olan 95–99 aralığının orta noktasıdır. yani 97'dir. Ayrıca formül kullanılarak ta mod'u bulmak mümkündür.

$$\text{Mod} = l + \left(\frac{f_{ü}}{f_{ü} + f_a}\right).\ a$$ formülünde:

$l =$ Frekansı en yüksek olan aralığın alt sınırı

$f_{ü} =$ Frekansı en yüksek olan aralığın üstünde kalan aralıktaki frekans

$f_a =$ Frekansı en yüksek olan aralığın altında kalan aralıktaki frekans, ve

$a =$ Aralık ölçüsüdür.

Şimdi değerleri yerleştirelim.

$$\text{Mod} = 95 + \left(\frac{50}{50 + 45}\right).\ 5$$

$$= 95 + 2.55$$
$$= 97.55$$

Görüldüğü gibi sadece dağılıma bakarak bulunan mod ile formül kullanmak sureti ile bulunan arasında büyük bir farklılık yoktur. Meydana gelen farklılık ise verilerin aralıklar için guruplanması ile ilgilidir.

Dağılımlarda birden fazla mod olabilir. İki ve daha fazla mod'u olan dağılıma "bimodal" adı verilir.

Özetle, mod bir vasat değeridir. Uçtaki gözlem değerlerinden etkilenmez, geniş bir örneklemden elde edilen dağılımlarda daha anlamlı bir merkezî eğilim ölçüsü olup, ileri istatistik işlemlere olanak vermez.

ii. Medyan (Ortanca)

Diğer basit bir merkezî eğilim ölçüsü de, ölçekte bir noktayı belirleyen medyan'dır. Bu noktanın altında gözlenen ölçümlerin yarısı ve üstünde ise diğer yarısı bulunmaktadır. Yedi kişilik bir öğrenci grubunun puvanları, 85, 80, 78, 76, 70, 67, 60 olsun. Burada N/2, diğer bir deyişle dağılımı tam eşit iki parçaya bölen nokta 76, medyandır. Görüldüğü gibi, dağılımdaki gözlem sayısı tek sayılı olduğu zaman bu kolaydır. Çift olduğu takdirde dağılımı iki eşit parçaya bölen aralığın orta noktasını almak gerekmektedir. Şimdi grubu 8 öğrenci olarak düşünelim ve puvanları, 85, 80, 78, 76, 70, 67, 60 ve 55 olsun. Burada medyan 70–76 aralığının orta noktası olan 73'dür.

Pratik yöntemle bulunan medyanı daha başka durumlara da uygulanabilecek nitelikte formüle etmek istersek;

$$\text{Medyan} = X_a + \left(\frac{X_b - X_a}{2}\right)$$

şeklinde tanımlanır. Burada

X_a = Medyanın kendisinden büyük olduğu ölçümün gerçek üst sınırı, ve

X_b = Medyanın kendisinden küçük olduğu ölçümün gerçek alt sınırıdır.

8 kişilik öğrenci grubundan elde edilen puvanlara uygulayalım; Bu dağılımda;

$X_a = 70.5$
$X_b = 75.5$

olduğuna göre

$$\text{Medyan} = 70.5 + \left(\frac{75.5 - 70.5}{2}\right)$$

$$= 73 \text{ olmaktadır.}$$

Gruplanmış serilerde, doğal olarak, farklı bir formül uygulanmaktadır.

$$\text{Medyan} = l + \left(\frac{N/2 - tf_a}{f_i}\right).a$$

Burada:

l= Medyanın bulunduğu aralığın gerçek alt sınırı

tf_a= Medyanın bulunduğu aaralıktan önce gelen aralıklardaki toplam ferekans

f_i= Medyanın bulunduğu aralıktaki frekans, ve

a= Aralık ölçüsüdür.

Tablo: 1'deki dağılımın medyanını bulalım. Toplam frekans yani N= 270 dir.

N/2 135, l= 94.5, tf_a = 70, f_i= 70, a= 5 olduğuna göre;

$$\text{Medyan} = 94.5 + \left(\frac{135 - 70}{70}\right).\ 5$$

$$= 94.5 + (0.93).\ 5$$
$$= 94.5 + 4.65$$
$$= 99.15 \text{ olarak bulunur.}$$

Görüldüğü ve mod'da olduğu gibi medyan da dağılımda kopukluk gösteren uç gözlem değerlerinden etkilenmez. Bu yüzden de dağılımın bu tip gözlem değerlerden etkileneceği düşünülürse, merkezi eğilim ölçüsü olarak medyan tercih edilir. Medyanı hesaplamak kolaydır. Fakat, aritmetik ortalamadan daha az güvenilir ve ileri istatistiklere pek fazla olanak vermez.

iii. Aritmetik Ortalama

Bilindiği gibi, bir çok istatistiklere kaynak olan aritmetik ortalama, dağılımdaki gözlem değerlerinin toplamının, gözlem sayısına bölümü olarak tanımlanmaktadır.

Aritmetik ortalama şu formülle ifade edilir:

$$\bar{X} = \frac{\Sigma X}{N}$$

On kişilik bir öğrenci grubunun aldığı puvan dağılımını 60, 59, 58, 53, 47, 45, 45, 40 39, 36, olarak kabul edersek, aritmetik ortalama,

$$\bar{X} = \frac{482}{10}$$

$$= 48.2 \text{ dir.}$$

Gruplanmış bir seride ise ortalama hesaplanmasında en kolayı olması bakımından şu formül daha yaygın olarak kullanılmaktadır.

$$\bar{X} = \frac{\Sigma f X_o}{N}$$

Burada, f= Frekans, ve

X_o= Puvan aralığının orta noktasıdır.

$\varepsilon f X_o$= Puvan aralığının orta notktalarının frekans ile çarpımlarının toplam,

Tablo 1'de ki dağılımı yine kullanalım. Bu dağılımda

$\varepsilon f X_o$= 27390, ve

N= 270 olduğuna göre; ortalama

$$\bar{X} = \frac{27390}{270}$$

$$= 101.4 \text{ eder.}$$

Yukarıda yapılan işlem, her aralığın orta noktasını o aralıktaki frekansla çarpmak ve herbir aralık için ayni işlemi yaptıktan sonra toplamını almak sureti ile εfX_0 bulunmaktadır. Bulunan değer gözlem sayısına bölündükten sonra aritmetik ortalama hesaplanmış olur.

Aritmetik ortalama, en çok bilinen ve kullanılan bir merkezî eğilim ölçüsüdür. Kolaylıkla anlaşılır. Fakat mod ve medyandan farklı olarak, aritmetik ortalama dağılımda kopukluk gösteren uç gözlem değerlerinden etkilenir. Bu yüzden bu tip bir uç değer olduğu zaman, ortalamanın tipik bir merkezî eğilim ölçüsü olması şüphe ile karşılanabilir. Aritmetik ortalama daha ileri istatistik analizler için önemli bir kaynak olmaktadır.

iv. Ortalamaların Ortalaması

Bazı sorunlar için, örneğin ayni populasyondan seçilmiş birden fazla örneklemelerden elde edilen ortalamaların ortalamasını bulmak yararlı olmaktadır. Örneğin politik davranışlar konusunda seçilmiş üç örnekleme uygulanan bir ölçek sonucu;

Birinci örneklemin $n_1 = 100$ $\bar{X}_1 = 15.0$
İkinci örneklemin $n_2 = 80$ $\bar{X}_2 = 12.0$, ve
Üçüncü örneklemin $n_3 = 60$ $X_3 = 16.0$ olsun

Ortalamaların ortalamasını veren formül

$$\bar{\bar{X}} = \frac{\Sigma n_j \bar{X}_j}{N} \qquad (J = \text{ortalama sayısı})$$

olduğuna ve bizim üç örneklerimiz bulunduğuna göre

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2 + n_3 \bar{X}_3}{N} \text{ olmaktadır.}$$

Burada

$N = n_1 + n_2 + n_3$ tür.

Buna göre, yukardaki örnekte ortalamaların ortalaması;

$$\bar{X} = \frac{(100 \times 15) + (80 \times 12) + (60 \times 16)}{100 + 80 + 60}$$

$$= \frac{1500 + 960 + 960}{240}$$

$$= \frac{3420}{240}$$

$= 14.2$ olmaktadır.

Görüldüğü gibi, üç örneklemden elde edilen ortalamalar tek bir ortalama puvanla ifade edilebilmektedir.

c. Yayılım Ölçüleri (Değişim Öüçüleri)

Elde edilen verilerin merkezî eğilim ölçülerini bilmek dağılımın özelliklerini bilme bakımından tam bir bilgi vermemektedir. Örneğin, iki dağılımın ortalamasının ayni olması yeter bir bilgi değildir. Bu ortalamanın altında ve üstünde kalan gözlem değerlerinin birbirlerinden olan farklılıklarını bilmek gerekir. Aksi halde yetersiz yorumlar yapılabilir. Bu yüzden gözlem değerlerinin nasıl bir yayılım gösterdiğini bilmek istatistiğin bir diğer ilgi merkezidir. Burada en basitten başlamak üzere ranj, ortalama kayma ve standart kayma kavramlarına yer verilecektir.

i. Ranj

Ranj, kısaca bir dağılımda en alt gözlem değer ile en üst gözlem değer arasındaki açıklığa denmektedir. Örneğin, bir dağılımda en az gözlem değer 20 ve üstteki 80 ise ranj 80 – 20= 60 olmaktadır. Gruplanmış serilerde ise en alttaki aralığın orta noktası ile en üstteki aralığın orta noktası arasındaki fark alınır. Tablo 1'deki dağılımda en alt aralıktaki orta nokta 77 ve en üstteki 122 olduğuna göre ranj 122–77= 45 olmaktadır. Ranj çok basit bir yayılım ölçüsü olarak fazla yoruma olanak vermez.

ii. Ortalama Kayma

Bir dağılımın ortalama kayması, o dağılımdaki gözlem değerlerinin herbirinin, dağılımın ortalamasından olan mutlak uzaklıklarının toplamının gözlem sayısaına bölünmesi ile bulunur.

Formülü

$$O.K = \frac{\Sigma\,|x|}{n}$$ dir. Gruplanmış seri için ise formüle bir f eklenir.

$$O.K = \frac{\Sigma\, f\,|x|}{n}$$

Her iki formüldeki $|x|$ ise
$|x| = |X - \bar{X}|$ demektir.

On kişilik bir gruba verilen ölçek sonucu alınan puvanlar 22, 24, 26, 30, 32, 35, 37, 39, 41, 44 olsun. Bu örnek için ortalama kaymayı bulmak için önce aritmetik ortalamanın bulunması gerekir. Bu serinin aritmetik ortalaması 33 tür. Her gözlem değerinden çıkarıldığı zaman, 33–22=11 ve diğerleri de 9, 7, 3, 1, 2, 4, 6, 8 ve 11 olmaktadır.

Şimdi, formüle göre

$$O.K = \frac{64}{10}$$

$= 6.4$ bulunmuş olmaktadır.

Ortalama kayma, Sosyal Bilimlerde nadir uygulanan bir yayılım ölçüsüdür. Fakat bundan sonra sözü edilecek standart kayma kavramının daha kolay anlaşılması bakımından burada kısaca söz konusu edilmesinde yarar bulunmuştur.

iii. Standart Kayma

Bir dağılımın gösterdiği değişimin en güvenilir ölçüsü standart kayma'dır. Bir dağılımın standart kayması ise, o dağılımdaki gözlem değerlerinin ortalamadan olan farklarının kareleri ortalamasının kare köküne eşit olmaktadır.

Standart kayma böylece;

$$s = \sqrt{\frac{\Sigma x^2}{N}}$$

olarak dormüle edilir. Yukarda, ortalama kayma için verilen bilgilerin standart kaymasını bulalım. Önce Tablo 3'de standart kayma için gruplanmamış bir seride gerekli işlemleri gösterelim.

TABLO 3.
Gruplanmamış bir seri için
Standart kaymaye ait Bilgiler

(Puan) ×	(Ortalamadan farklılık $X - \underline{X} = X$	(Farklılığın karesi) $(X - \underline{X})^2$
22	— 11	121
24	— 9	81
26	— 7	49
30	— 3	9
32	— 1	1
35	+ 2	4
37	+ 4	16
39	+ 6	36
41	+ 8	64
44	+ 11	121
Σ	0	502

Değerlerin formüldeki yerlerine koyalım

$$s = \sqrt{\frac{502}{10}}$$

$$= \sqrt{50.2}$$

$$= 7.1$$

Gruplanmış serilerde standart kayma için yine çeşitli yöntemleri uygulayan formüller düşünülebilir. Bunlardan en yaygın olanı, çıkarma ve bölme işlemlerini beraberce uygulayarak kodlama yolu ile standart kayma hesaplanmasıdır.

Şöyle formüle edilmektedir:

$$s = a.\sqrt{\frac{\Sigma fx^2}{n} - \left(\frac{\Sigma fx'}{n}\right)^2}$$

Şimdi Tablo 1'deki bilgileri bu formüle göre Tablo 4'de gösterelim.

TABLO 4.

Gruplanmış bir seri için standart kayma

x	f	Xo	x'	fx'	fx'¹
120 - 124	5	122	9	45	405
115 - 119	10	117	8	80	640
110 - 114	25	112	7	175	1225
105 - 109	40	107	6	240	1440
100 - 104	50	102	5	250	1250
95 - 99	70	97	4	280	1120
90 - 94	45	92	3	135	405
85 - 89	15	87	2	30	60
80 - 84	5	82	1	5	5
75 - 79	5	77	0	0	0
Σ	270			1240	6550

Burada, çıkarma ve bölme işlemleri X_0 ve x' kolonlarında yapılmaktadır. Çıkarılan sabit sayı en alt aralığın orta noktası ve bölünen sabit sayı ise aralık ölçüsüdür. Bir örnek yapalım:

En üst aralığın orta noktası 122 olduğuna göre, $122-77=45\div5=9$'dur. x' kolonuna 9 yazılmıştır. Diğerleri de buna göre hesaplanır. Görüldüğü gibi böylece işleme girecek sayılar çok küçültülerek standart kayma hesaplanmasındaki işlemler kolaylaştırılmış olmaktadır.

Şimdi, bulunan değerleri formüldeki yerlerine koyalım:

$a=5$ $N=270$ $\Sigma fx = 1250$ ve $\Sigma fx^2 = 6550$ olduğuna göre

$$s = 5 \sqrt{\frac{6550}{270} - \left(\frac{1240}{270}\right)^2}$$

$$= 5 \sqrt{24.2 - 21.16}$$

$$= 5 \times 1.75$$

$= 8.75$ olarak bulunur.

Bir yayılım ölçüsü olan standart kayma gözlem değerlerinin dağılımı ile doğrudan ilişkilidir. Bir dağılımda gözlem değerlerinin birbirinden olan farklılıkları arttıkça standart kayma büyür, azaldıkça standart kayma küçülür. Böylece ölçme yapılan grubun homojen ya da heterojen oluşu hakkında bir yargıya varmada standart kayma yardımcı olur.

d. Vardamlı istatistikten Bazı Teknikler

Bilindiği gibi, Sosyal Bilim araştırmalarının hemen hemen hepsi, araştırma amaçları yönündeki verileri populasyondan seçilmiş örneklem ya da örneklemlerden toplamaktadırlar. Çünkü araştırma konusunu populasyonun tümünde gözlemek birçok bakımdan imkânsız ve çoğu kerede gereksizdir. O halde, araştırıcılar örneklemde elde edilen istatistikleri populasyon için genellemek durumundadır. İşte istatistiğin bu yöndeki uğraşısı "vardamlı istatistik" dediğimiz bir türü ortaya çıkarmıştır. Buraya kadar bazı örnekleri kısaca tartışılan istatistik teknikler ise doğrudan örneklemin özelliklerini tanıtan "betimsel istatistik"in konuları arasındadır.

Bir örneklemden elde edilen istatistikleri populasyon için genelleme yönünde çok çeşitli ve ayrıntılı teknikler vardır. Bunların hepsini burada vermek mümkün olmadığı gibi gerekli de değildir. Yalnız konumuzla ilgili olduğu düşünülen t– testi, F– testi'nin tanıtılması yararlı olacağı düşüncesi ile bu iki önemli tekniğe kısaca yer verilecektir.

i. t– Testi

Sosyal Bilim araştırmalarında aritmetik ortalama oldukça yaygın olarak kullanılan bir istatistiktir. Çoğu kere de iki bağımsız örneklemden elde edilen iki ortalamanın birbirinden farklılıklarının gerçekten anlamlı mı yoksa tesadüflere bağlı mı olduğu sorunu ortaya çıkmaktadır. Bu sorunun çözümünde t– testi büyük ölçüde yardımcı olmaktadır.

t– testi şöyle tanımlanmaktadır.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S\bar{x}_1 - \bar{x}_1}$$

Burada,

$\bar{X}_1$ = Birinci örneklemden elde edilen aritmetik ortalama
$\bar{X}_2$ = İkinci örneklemden elde edilen aritmetik ortalama, ve
$S\bar{x}_1 - \bar{x}_1$ = Farklılığın standart hatasıdır.

Farklılığın standart hatasını bulmak için;

$$S\bar{x}_1 - \bar{x}_2 = \sqrt{\left(\frac{\Sigma\bar{x}_1 + \Sigma\bar{x}_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

formülü uygulanır. Fakat burada Σx^2'nin bulunmasında da bir diğer formülden yararlanpak gerekmektedir,

$$\Sigma x_1^2 = \Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n_1}$$

Şimdi bir örnekte bunu kullanalım. İki ayrı örnekleme bir ölçek uygulanmış ve Tablo 5'deki sonuçlar alınmış olsun:

TABLO 5.

İki örneklerden alınan gözlem değerleri

Örneklem 1		Örneklem 2	
X_1	X_1^2	X^2	X^2_1
15	225	10	100
16	256	12	124
19	361	14	196
19	361	16	256
19	361	17	289
20	400	19	361
21	441	20	400
26	676	22	484
27	729	23	529
35	1225	25	625
Σ 217	5035	178	3364

t– testi için, örneklemlerden elde edilen dağılımların normal olduğunu kabul etmek gerekmektedir. Biz de örneğimizde normal dağılım elde ettiğimiz kabul ederek işlemlere devam edelim.

Önce, örneklem 1 için

$$\Sigma x_1^2 = 5035 - \frac{47089}{10}$$

$$= 5035 - 4709$$

$$= 326$$

Sonra örneklem 2 için

$$\Sigma x_2^2 = 3364 - \frac{31684}{10}$$

$$= 3364 - 3168$$

$$= 196$$

Şimdi εx_1^2 ve εx_2^2 değerlerini farklılığın standart hatası formülünde yerlerine koyalım;

$$S\bar{x}_1 - \bar{x}_2 = \sqrt{\left(\frac{326 + 196}{18}\right)(0.2)}$$

$$= \sqrt{\left(\frac{522}{18}\right)(0.2)}$$

$$= \sqrt{5.8}$$

$$= 2.4$$

Son olarak bulunan farkların standart hatasını t-testi formülüne yerleştirmeden önce $\bar{X}_1$ ve $\bar{X}_2$'yi saptayalım.

$\bar{X}_1 = 20.5$ ve $\bar{X}_2 = 17.8$'tir.

Böylece;

$$t = \frac{20.5 - 17.8}{2.4}$$

$$= \frac{2.7}{2.4}$$

$= 1.13$ çıkmaktadır.

Yüzde 5 manidarlık seviyesi ve $n_1 + n_2 - 2 = 18$ serbestlik derecesine göre t'nin teorik–tablo değeri 2. 101 olarak görünmektedir. t'nin teorik–tablo değeri, örneklemlerimiz için hesaplanan t–değerinden büyüktür. Böylece, söz konusu örneklemlerimizin ortalamaları arasındaki farklılık anlamlı ya da tesadüfen meydana gelmeyen bir farklılık olarak nitelendirilemez.

ii. F–Oranı ve F–Testi

t–testi yolu ile iki ortalama arasındaki farklılığın değerlendirilmesinde, her iki örneklemin varlığının birbirine eşit olduğunu düşüncesinden hareket edilebilir. Diğer yandan bu çoğu kere doğru değildir. Yani varyanslar da ortalamalarda olduğu gibi bazı farklılıklar ortaya koyacaktır. İşte örneklemlerin varyansları arasındaki farklılığın tesadüfler dışı olduğunu ve bunun derecesini F–oranı ve F–testi ile görmek mümkündür.

Tablo 5'deki bilgilere göre, birinci populasyonun varyansının, ikinci populasyonun varyansına eşit olduğu ($\sigma_1^2, = \sigma_2^2$) hipotezini kabul veya red edebilmek için F–oranını ve F–testini uygulayalım. Bu teknikleri kullanırken hereket noktamız şudur; örneklemlerin alındıkları populasyonların varyanslarının ve dolayısı ile örneklemlerin varyanslarının farklı olmadığını, diğer bir deyişle, eşit olduğunu öneren hipotezi, hipotetik örneklemlerde elde ettiğimiz değerlerle karşılaştırmaktır.

F–oranı formülü

$$F = \frac{S_1^2}{2_2^2} \quad \text{veya} \quad F = \frac{S_2^2}{S_1^2} \quad \text{dir.}$$

Burda

$$S_1^2 = \frac{\Sigma x_1^2}{n_1 - 1} \quad \text{ve}$$

$$S_2^2 = \frac{\Sigma x_2^2}{n_2 - 1}$$

olarak tanımlanmaktadır.

Örneklem 1 için: $\Sigma x_1^2 = 326$

Örneklem 2 için: $\Sigma x_1^2 = 196$ olarak bulunmuştur.

Buna göre

$$S^2{}_1 = \frac{326}{10-1}$$

$$= 36.2$$

$$S^2{}_2 = \frac{196}{10-1}$$

$$= 21.8$$

F–oranını bulmak için, bulunan yüksek s değerini paya koymak sureti ile;

$$F = \frac{36.2}{21.8} \text{ elde edilir, ve}$$

sonuç olarak F= 1.62 çıkar.

Bulunan F–oranının tablodaki teorik oranları ile karşılaştırmak için F–testini uygulamak gerekir.

F–tablosuna gidebilmek için serbestlik derecesi bulunur. O da:

$n_1 - 1 = 10{-}1 = 9$, ve
$n_2 - 1 = 10{-}1 = 9$, dur.

Şimdi 9 ve 9 olan serbestlik derecelerine göre F tablosuna bakıldığında yüzden 5 manidarlık seviyesinde F değerinin 3.18 olduğu görülür. Buna göre Tablo 5'teki bilgilerin ortaya koyduğu iki örneklemde, varyanslar arasında manidar bir farklılık gözlenmemektedir. Diğer bir değişle birinci populasyonun varyansının, ikinci populasyonun varyansına eşit olduğu ($\sigma^2 = \sigma^2_2$) hipotezi doğrulanmamış olmaktadır.

Bu kısımda söz konusu edilen t–testi ve F–oranı ile F–testini uygulamada çeşitli koşulları gözününde bulundurmak gerekmektedir. Bu yüzden analizde söz konusu edilen verilerin türüne ve sağlanan koşullara göre en uygun gelen manidarlık testini saptayabilmek için daha ayrıntılı bilgilere ihtiyaç vardır. Bu yüzden, bu konularda gerekli ayrıntılı bilgiler istatistik kaynak kitaplarından edinilmelidir. Çünkü; burada bu kavramlar sadece genel hatları ile analiz teknikleri açısından hatırlatılmaktadır.

e. Kay–Kare (x^2)

Sosyal Bilim araştırmalarında, sık sık, obje, tepki ve özellik sayıları ilgi merkezi olabilir. Örneğin, politik davranışlarla ilgili bir araştırmada, kırsal ve kentsel bölgelerde oturanların ne kadarının ilk seçimlerde oy vermek ya da vermemek düşüncesinde olduğunu bilmek önemli olabilir. Bundan önce kısaca tartışılan t–testi ve F–oranı ile F–tesinde, dikkat edildiği gibi, örneklemlerden elde edilen ortalama ve varyansdan populasyona ait ortalama ve varyans tahmin edilmektedir. Diğer bir deyişle, örneğin, örneklemden elde edilen bir ortalamanın, hipotetik ortalama (beklenen ortalama) dan olan farkı aranmaktadır.

Benzer bir yaklaşım non– parametrik bir istatistik olan kay–kare için söz konusudur. Bunda aranan, elde edilen frekans ile beklenen ya da hipotetik frekans arasındaki farklılıktır. Bu farklılık ya şansa dayanan ya da belli bir güvenirlik sınırı içinde gerçek bir farklılığı yansıtır.

Bir örnek üzerinde durumu görelim. Kırsal bölgelere göre oy verip- vermeme eğilimi sorunu şöyle bir tablo vermiş olsun;

Tablo 6.

Kırsal bölgelerde oy verme oy verme eğilimine ait sayılar

	Oy verecek	Oy vermeyecek	Toplam
Gözlem	57	83	140
Beklenen	70	70	140

Tablo 6'da elde edilen (gözlenen) frekanslar verilmiştir. Şimdi beklenen frekansı (hipotetik frekans – teorik frekans) bulmak gerekir. Şöyle bir yol tutulabilir. "Kırsal bölgede oy vermek isteyenlerin sayıları istemeyenlere eşittir." hipotezinden hareketle örneğin, kırsal bölgelerde örnekleme giren 140 kişinin 70'i oy verecek ve 70'i oy vermeyecek şekilde beklenen frekanslar gösterilebilir.

Buna göre kay – kare; X^2

$$x^2 = \Sigma \frac{(f_g - f_b)^2)}{f_b} \quad \text{olarak tanımlanır.}$$

Burada

$f_g =$ Gözlenen frekans, ve

$f_b =$ Beklenen frekanstır.

Tablo 5'deki sayılar uygulanırsa,

$$x^2 = \frac{(57 - 70)^2}{70} + \frac{(83 - 70)^2}{70}$$

$$= 2.41 + 2.41$$

$$= 4.82 \text{ bulunmaktadır.}$$

Tablo değerine gitmek için serbestlik derecesi;

$$(n - 1)(k - 1)$$

formülünden elde edilir. Burada;

n= Sıra sayısı

k= Sütun sayısıdır.

Tablo 6'da n= 2 ve k = 2 olduğuna göre, serbestlik derecesi (2 – 1) (2–1) = olmaktadır.

Yüzde 95 güvenirlik seviyesinde kay karenin 1 serbestlik derecesinde tablo değeri 3.841'dir. Tablo 6 için bulunan kay – kare değeri (4.82) tablo değerinden büyük olduğu için, kırsal bölgelere oy vermeyecek olanların, oy verecek olanlardan daha fazla sayıda olduklarına gözlemin tesedüf dışı olduğu yorumu yapılabilir. Ancak yüzde 95 ihtimalle bu sonucun, farklılığının olın tesadüf dışı olabileceği düşünülmelidir.

Ayni hipotetik örnekte, diğer bir değişkeni daha dikkate alalım. Bu sefer oy verip, vermeme eğilimini kırsal ve kentsel bölgelere göre düşünelim ve Tablo 7'nin bir araştırma sonucu elde edildiğini kabul edelim.

TABLO 7.

Kırsal ve Kentsel Bölgelere göre oy verme eğilimi

	Oy verecek		Oy vermeyecek		Toplam
	Gözlenen	Beklenen	Gözlenen	Beklenen	
Kırsal	57	46	83	94	140
Kentsel	25	36	85	74	110
Toplam	82	82	168	168	250

Burada beklenen frekanslar, gözlenen frekansın bulunduğu sıra toplamı, sütun toplamı ile çarpılarak, toplam frekansa bölünmek sureti ile saptanır. Örneğin, gözlenen frekans 57 için,

$$\frac{82 \text{ x } 140}{250} = 46$$

beklenen frekanstır. Diğerleri de buna göre hesaplanarak Tablo 7'de ilgili yerlere konmuştur.

Şimdi kay – kare formülüne göre sonuç,

$$x^2 = \frac{(57-46)^2}{46} + \frac{(83-94)^2}{94} + \frac{(25-36)^2}{36} + \frac{(85-74)^2}{74}$$

$= 2.63 + 1.29 + 3.38 + 1.63$

$= 8.93$ tür.

Tabloya gidebilmek için serbestlik derecesi;

(2 – 1) (2 – 1) = 1 olarak bulunur. Kay – kare değeri ise, yüzde 1 güvenirlik seviyesinde 6.635 olduğuna göre, bölgeler ile oy

verme arasında çok yakın bir ilişki olduğu görülebilmektedir. Diğer bir deyişle kentsel bölgelerde oturanlar, kırsal bölgelerde oturanlara göre daha az sayıda oy vereceklerdir. Bu yorum ise yüzde 99 güvenirlik sınırı içinde yapılabilmektedir. Pratik bir açıklama ile, eğer bu araştırma yüz defa yapılırsa ancak bir seferinde bu ilişki bulunmayabilir denilmektedir.

Kay– kare iki değişken arasında ilişkiyi görmek için kullanıldığı gibi ikiden fazla değişkenler arasındaki ilişki için de düşünülebilir. Bu takdirde uygulanack yöntemler için istatistik kitaplarına bakılmalıdır.

Kay– kare, görüldüğü gibi, esasta iki değişken arasındaki ilişkiyi göstermektedir. Bu ilişkinin miktarı hakkında bilgi vermemektedir. Eğer böyle bir ihtiyaç hissedilirse o takdirde bazı teknikler kullanılabilir. Bunlardan en yaygın olanı kay – kareden yararlanarak hesaplanan kontincensi katsayısıdır (contingency coefficient).

Kontincensi Katsayısı;

$$C = \sqrt{\frac{x^2}{n + x^2}} \quad \text{olarak tanımlanır.}$$

Tablo 7'deki veriler için hesaplanan kay–kareye ($x^2 = 8.93$) göre kontincensi katsayısı;

$$C = \sqrt{\frac{8.93}{240 + 8.93}}$$

$$= \sqrt{\frac{8.93}{258.93}}$$

$$= \sqrt{0.034}$$

$$= 0.19 \text{ olarak bulunur*.}$$

Eğer elde, Tablo 7'de olduğu gibi her iki değişkenin iki alt değişkene bölünme durumu varsa, bu takdirde bir başka yolla ilişki ölçüsünü bulmak mümkündür. Yule katsayısı adı verilen bu ilişki ölçüsüsünün hesaplanmasına bir örnek verelim.

---

*Elde edilen C'nin yorumunda, bunu elde edilebilecek en yüksek C'ile karşılaştırmak gerekir. En yüksek C'yi bulmak için, Cmax = $\sqrt{1 - \frac{1}{n}}$ — formülü kullanılır. Burada n = kategori sayısıdır.

Aşağıdaki şekilde görüldüğü gibi tablo elde edilmiş olsun.

A	B
C	D

Yule katsayısı ise,

$$Q = \frac{AD - BC}{AD + BC} \text{ olarak tanımlanmaktadır.}$$

Örneğin, bir araştırma sonucunda;

	Oy verecekler	Oy vermeyecekler
Kırsal	20	30
Kentsel	15	40

elde edilmişse

AD = 800

BC = 450 olduğuna göre

$$Q = \frac{800 - 450}{800 + 450}$$

$$= \frac{350}{1250}$$

= 0.28 olarak bulunur.

Görüldüğü gibi Yule katsayısında, frekansların bulunduğu hücrelerin yeniden düzenlenmesi öngürülmektedir. Eğer AD = BC olursa o takdirde hiç bir ilişki görülmüyor ya da, AD veya BC'nin biri O ise, o takdirde tam bir ilişki bulunacaktır.

Buraya kadar, analizde kullanılabilecek istatistik teknikler yönünde çok yaygın olan bazı örnekler verilmiştir. Diğer yandan, Sosyal Bilim araştırma konuları oldukça karmaşıktır. Araştırılan konuların karmaşıklığı oranında da, etkin sonuçlara gidebilmek için, istatistik tekniklerin çok daha titizlikle seçilmesi gerekir. Bu yüzden, Sosyal Bilimlerde araştırıcılar, analizde kullanılacak istatistik teknikler için, daha çeşitli yöntemleri ayrıntılarına inen bir şekilde veren istatistik kitaplarına sık sık başvurmak durumundadırlar.

YARARLANILAN KAYNAKLAR

Arıcı, Hüsnü. *İstatistik - Yöntemler ve Uygulama* - Hacettepe Üıniversitesi Basımevi. Ankara, 1972.

Arkın, Herbert, and Colton, R, *Ekonomi, İşletmecilik, Psikoloji, Eğitim ve Biyolojiye Uygulanan İstatistik Metotlar* (Çeviren, Saim Kendir), Ankara, 1968.

Ayatar, Hazım. *Uygulamalı Eğitim İstatisliği.* Milli Eğitim Bakanlığı Eğitim Birimi, Ankara, 1971.

Edwards, Allen L., *Statistical Methods for the Behavioral Sciences,* Rinehart and Company, İnc., New York, 1956.

Galtung, Jopan., *Theory and Methods of Social Research,* Columbia University Press, New York, 1969.

Glock, Charles Y., *Survey Research in the Social Sciences,* Russell Sage Foundation, New York, 1967.

Good, Carter R., Scates, Douglas E, *Methods of Research,* Appleton Century Crofts, New York, 1960.

Gopal, M.H., *An Introduction to Research Procedure in Social Sciences,* Asia Publishing House, New York, 1970.

Goode, J. William and Hatt, Paul., *Methods in Social Research,* Mc Graw-Hill Book Company, Inc, New York, 1952.

Hymen, Herbert., *Survey Design and Analysis,* the Free Press, Publishers, Glencoe, İllinois, 1960.

Lindquist, E.F., (Çevirenler: Tan, Hasan ve Taner, Tuğrul)., *İstatistiğe Giriş,* Milli Eğitim Bakanlığı Öğretmen Kitapları No: 90, İstanbul, 1968.

Selltiz, Claire, Marie., Jahoda, Deutsch, Morton, Cook, Stuart W., *Research Methods in Sociaı Relations,* Holt, Rinehart and Winston, New York, 1959.

Walker, Helen M. and Lev, Joseph., *Statistical İnference,* Henry Holt and Company, New York, 1953.

Webb, Eugene J., Campbell, Denald T., Schwartz, Ricgard D., and Sechrest, Lee., *Unobtrusive Measures- Non reative Research in the Social Sciences,* Rand McNally and Co. Chicago, 1971.

Young, Pauline V., *Scientific Social Surveys and Research,* Prentice-Hall of İndia Private Limited, New Delhi, 1968.