

BÖLÜM 1: FREKANS DAĞILIMLARI

1.1. Giriş

İstatistik, rasgelelik içeren olaylar, süreçler, sistemler hakkında modeller kurmada, gözlemlere dayanarak bu modellerin geçerliliğini sınamada ve bu modellerden sonuç çıkarmada gerekli bazı bilgi ve yöntemleri sağlayan bir bilim dalıdır.

İstatistiksel tekniklerin biyoloji alanında istatistikle ilgili olan problemlerin çözümüne uygulanması biyoistatistiğin konusudur.

1.2. Bazı Temel Kavramlar

Kitle: Belirli bir özelliğe sahip bireylerin veya birimlerin tümünün oluşturduğu topluluğa **kitle** denir.

Örneklem: Örnekleme yöntemlerinden yararlanılarak bir kitleden seçilen, aynı özellikleri taşıyan ve kitleyi temsil edebilecek nitelikteki ve nicelikteki bireylerin oluşturduğu topluluğa **örneklem** denir. Üzerinde araştırma yapılan kitle bazen sayılamayacak kadar çok birim içerebilir. Bu durumda bu birimlerin tamsayımı mümkün veya gerekli olmayabilir. Mümkün olsa bile zaman ve maliyet gibi kısıtlayıcılar tam sayımı imkansız kılabilir. Bu gibi durumlarda kitleden örneklem alınarak, elde edilen sonuçlar ile kitle için tahmin yapılabilir.

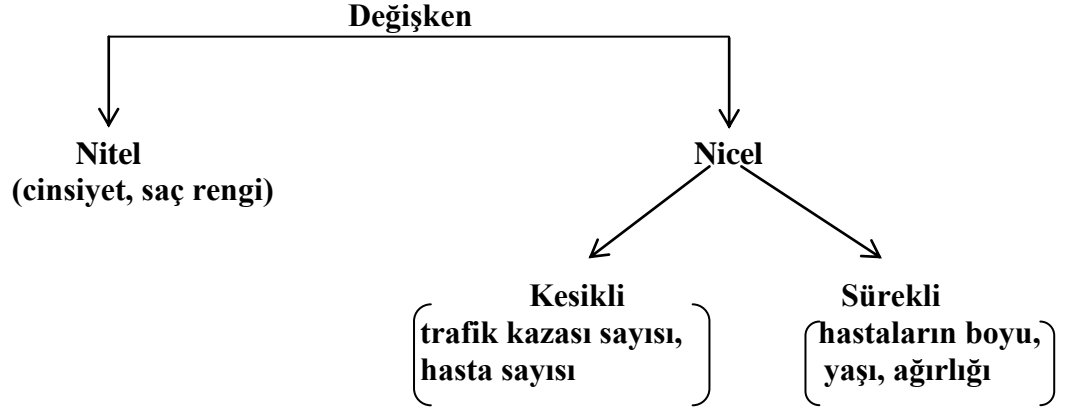
Parametre: Kitleyi tanımlayan sayısal değerlere **parametre** denir. Kitle ortalaması, kitle varyansı, kitle korelasyon katsayısı parametreye örnek olarak verilebilir.

Veri: Gözlemlerden elde edilen sayısal olan ya da olmayan sonuçlara **veri** denir.

Değişken: Birimlerin farklı değerler alabildikleri nitelik ya da niceliklerine **değişken** denir.

Bir değişken sayısal değerlerle ölçülebiliyorsa, bu değişkene nicel değişken denir. Örneğin ağırlık, boy uzunluğu, bir hastalığın iyileşme süresi gibi. Nicel değişkenler, kesikli ve sürekli değişkenler olarak ikiye ayrılır. Belli bir aralıkta her değeri alan değişkenler sürekli değişkenler, her değeri alamayan değişkenler kesikli değişkenler olarak adlandırılır. Örneğin hane halkı sayısı kesikli, hanedeki kişilerin ağırlıkları sürekli değişkenlerdir.

Bir deęişken sayısal deęerlerle ölçülemiyorsa, bu deęişkene nitel deęişken denir. Bunlara kategorik deęişken adı da verilir. Örneęin göz rengi, medeni durum, kan grubu gibi.



Birimlerden bilgi toplanmasında deęişkenlerin ölçülmesi önemli bir konudur. Deęişkenlerin ölçülmesi, genel olarak dört başlık altında açıklanabilir;

Sınıflama Ölçme Düzeyi: Birimlere özelliklerine göre isimler verilir. Ölçme, isimlendirilerek (sınıflandırılarak) yapılır, herhangi bir sıralama yapılmaz. Örneęin cinsiyet deęişkeni kadın, erkek gibi iki sınıfta ifade edilir. Bu çerçevede medeni durum, meslek, doğum yeri gibi deęişkenlere ait ölçümler için sınıflama ölçme düzeyi kullanılır.

Sıralama Ölçme Düzeyi: Sıralama ölçme düzeyinde deęişkenlerin aldığı deęerler önem derecesi ya da üstünlüklerine göre sıralanır. Katılım düzeyi (Kesinlikle Katılıyorum, Katılıyorum, Kararsızım, Katılmıyorum, Kesinlikle Katılmıyorum), sıklık düzeyi (Hiç, Nadiren, Genellikle, Her Zaman), öğrenim durumu (İlköğretim, Lise, Lisans, Yüksek Lisans), vb. deęişkenler için sıralama ölçme düzeyi kullanılır.

Eşit Aralıklı Ölçme Düzeyi: Sıcaklık, başarı, performans gibi nicel deęişkenleri ölçmek için kullanılır. Bu ölçekte bir başlangıç noktası bulunmaz. Yani, “0” deęeri eşit aralıklı ölçme düzeyinde yokluk ifade etmez. Örneęin termometrede görülen “0 C” belirli bir anlam taşır.

Oranlama Ölçme Düzeyi: Aylık gelir, aęırlık, uzunluk, hız gibi deęişkenleri ölçmek için kullanılır. Bu ölçme düzeyinde başlangıç “0” noktasıdır. Oran ölçme düzeyinde yer alan “0 Kg” bir yokluk ifadesidir.

Sınıflama ve sıralama ölçme düzeyi nitel deęişkenler için, eşit aralıklı ve oranlama ölçme düzeyi nicel deęişkenler için kullanılır.

1.3. Frekans Tabloları

1.3.1. Nicel Verilerde Frekans Tabloları

Örnek 1.1: Bir doğumevinde bir ayda doğan 120 erkek bebeğin ağırlıkları Tablo 1.1 de verilmiştir:

Tablo 1.1: 120 erkek bebeğin ağırlıkları(kg)

3.45	3.35	3.80	3.40	3.50	3.30	3.50	3.45	3.40	3.10	3.10	3.25
3.20	3.00	3.40	3.45	3.10	3.30	3.60	3.60	3.25	3.20	3.30	3.50
3.20	3.40	3.30	3.50	3.50	3.85	3.50	3.10	3.55	3.00	3.30	3.20
3.60	3.25	3.40	3.40	3.45	3.45	3.10	3.30	3.65	3.50	3.75	3.70
3.80	3.55	3.15	3.35	3.20	3.40	3.45	3.15	3.20	3.40	3.40	3.85
3.20	3.50	3.20	3.50	3.45	3.65	3.30	3.40	3.60	3.10	3.30	3.65
3.60	3.60	3.50	3.70	3.10	3.20	3.35	3.70	3.60	3.65	3.35	3.20
3.40	3.35	3.40	3.30	3.65	3.40	3.55	3.30	3.60	3.60	3.40	3.20
3.35	3.35	3.80	3.85	3.20	3.60	3.30	2.85	3.40	3.30	3.10	3.30
3.75	3.25	3.10	3.25	3.35	3.55	3.20	3.93	3.45	3.50	3.30	3.40

Örneği çözmeye başlamadan önce, frekans tablosunu oluşturabilmek için bazı tanımlar verilecektir.

Sınıf: Değişkenin değer aralığı birbirinden kesin olarak ayrılmış gruplara bölünebilir, bu gruplara sınıf adı verilir. Frekans dağılım tablosu hazırlamanın ilk adımı frekans dağılım tablosunda kaç sınıfın olacağına karar vermektir. Genellikle frekans dağılım tablolarında 8-10 sınıf yapılır. Sınıf sayısının 8'den az ve 15'den fazla olmamasına dikkat edilir. Eğer sınıf sayısı 15'i geçerse özetlemeden beklenen amaca ulaşılmamış olur. Sınıf sayısı k ile gösterilir. Sınıf sayısının belirlenmesinde aşağıda verilen Sturges formülü de kullanılabilir:

$$k = 1 + 3.3 \log(n)$$

Formülde n veri sayısıdır.

Dağılım Sınırları: Örneklemdeki en küçük değer ile en büyük değere dağılım sınırları denir.

Dağılım Genişliği: Bir örnekte en büyük değer ile en küçük değer arasındaki farka dağılım genişliği denir, R ile gösterilir.

Sınıf Aralığı: İki sınıf arasındaki farka sınıf aralığı denir, c ile gösterilir.

$$c = \frac{\text{dağılım genişliği}(\text{son hanesine 1 eklenmiş hali})}{\text{sınıf sayısı}(k)}$$

Alt Sınır: Bir sınıfın en küçük değeridir, A_s ile gösterilir.

Üst Sınır: Bir sınıfın en büyük değeridir, \bar{U}_s ile gösterilir.

Sınıf Değeri: Bir sınıfın alt sınır ve üst sınır değerlerinin ortalaması sınıf değeridir, S_i ile gösterilir.

Frekans (Sıklık): Bir sınıfa düşen veri sayısına frekans denir, f ile gösterilir. Frekansların toplamı veri sayısına eşit olmalıdır. Yani, $\sum_{i=1}^n f_i = n$ dir.

Görelî Frekans (Frekans Yüzdesi): Her sınıfa düşen veri sayısının toplam veri sayısına oranına denir, p ile gösterilir. $p_i = \frac{f_i}{n}, i = 1, 2, \dots, k, \sum_{i=1}^k p_i = 1$

Tablo 1.1 ile verilen 120 erkek bebeğin ağırlıklarına ilişkin frekans tablosu oluşturulsun.

Sınıf sayısı olarak $k = 10$ alındı.

$$R = 3.93 - 2.85 = 1.08$$

Sınıf aralığı,

$$c = \frac{\text{dağılım genişliği}(\text{son hanesine 1 eklenmiş hali})}{\text{sınıf sayısı}(k)} = \frac{1.08 + 0.01}{10} = 0.109 \cong 0.11$$

Tablo 1.2: 120 Erkek Bebeğin Ağırlıklarına İlişkin Frekans Tablosu

A_s	\bar{U}_s	Sınıf (S_i)	İşaretle frekans	Frekans(f_i)	Görelî frekans(p_i)
2.85	2.95	2.90	/	1	0.01
2.96	3.06	3.01	//	2	0.02
3.07	3.17	3.12	//// // /	11	0.09
3.18	3.28	3.23	//// // // //	18	0.15
3.29	3.39	3.34	//// // // // //	22	0.18
3.40	3.50	3.45	//// // // // //	35	0.29
3.51	3.61	3.56	//// // // //	14	0.12
3.62	3.72	3.67	//// //	8	0.07
3.73	3.83	3.78	////	5	0.04
3.84	3.94	3.89	////	4	0.03

Birikimli (Eklemeli) Frekans Tablosu

Her sınıfın üst sınırı ile bir sonraki sınıfın alt sınırı değerlerinin ortalaması sınıf ara değeri (*Sad*) olarak alınırsa, bu değerden daha az değer gösteren verilerin toplamı frekans sütununa yazıldığında den daha az frekans bulunmuş olur. Aynı şekilde, sınıf ara değerinden daha çok değer gösteren verilerin toplamı frekans sütununa yazıldığında den daha çok frekans bulunmuş olur.

Sınıf ara değeri, den daha az ve den daha çok sütunları tamamlandıktan sonra oluşan tabloya birikimli frekans tablosu denir.

Tablo 1.2: 120 Erkek Bebeğin Ağırlıklarına İlişkin Birikimli Frekans Tablosu

As	Üs	Sad	Den Daha Az		Den Daha Çok	
			fi	pi	fi	pi
		2.845	0	0.00	120	1.00
2.85	2.95	2.955	1	0.01	119	0.99
2.96	3.06	3.065	3	0.03	117	0.98
3.07	3.17	3.175	14	0.12	106	0.88
3.18	3.28	3.285	32	0.27	88	0.73
3.29	3.39	3.395	54	0.45	66	0.55
3.40	3.50	3.505	89	0.74	31	0.26
3.51	3.61	3.615	103	0.86	17	0.14
3.62	3.72	3.725	111	0.93	9	0.08
3.73	3.83	3.835	116	0.97	4	0.03
3.84	3.94	3.945	120	1	0	0.00

Birikimli frekans tablosuna bakılarak, bir ayda doğan 120 erkek bebeğin yüzde 45' i, 3.395 kilonun altındadır veya yüzde 73' ün 3.285 kilonun üzerindedir.

1.3.2. Nitel Veriler için Frekans Tablosu

Nitel veriler için sınıflama ve sıralama ölçme düzeyinin kullanılabilceği daha önce ifade edilmişti. Sınıflanabilen ve sıralanabilen verilerde sınıflar birbirinden bağımsız olduğu için frekans tablosunda sadece sınıf, frekans ve görel frekans sütunları yer alır.

Örnek 1.2: Bir hastaneye müracaat eden hastaların 400'ü verem, 300'ü kalp, 200'ü kanser, 100'ü grip ve 200'ü diğer hastalık türüne sahiptir. Bu verilere ilişkin frekans tablosu Tablo 1.4 de verilmiştir.

Tablo 1.4: Hastalık türüne göre hasta sayısı

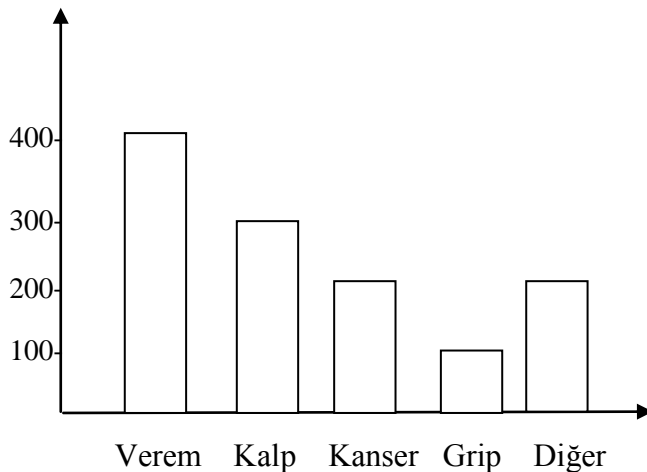
Sınıf(Hastalık Türü)	Frekans(Hasta Sayısı)	Görel Frekans
Verem	400	400/1200
Kalp	300	300/1200
Kanser	200	200/1200
Grip	100	100/1200
Diğer	200	200/1200

1.4. Grafikler

Grafikler verilerin çizgisel gösterimleridir. Grafikler frekans tablolarının tamamlayıcısı olarak düşünülebilir. Çok çeşitli grafik türleri vardır. Burada, çubuk grafiği, çizgi diyagramı, histogram(dağılım dikdörtgenleri), daire (pasta) grafiği verilecektir.

1.4.1. Çubuk Grafiği

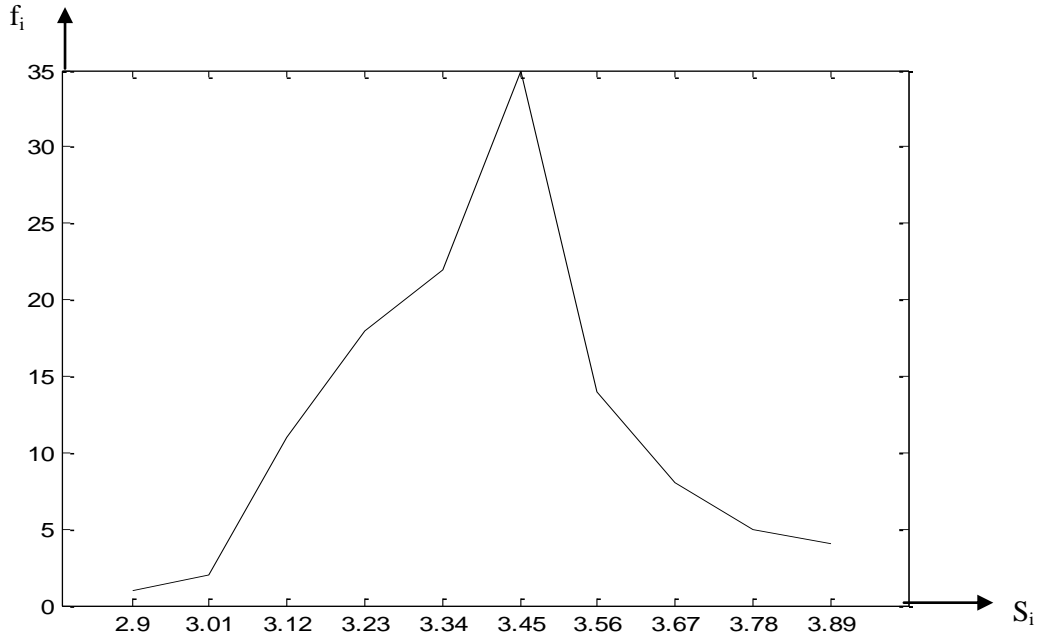
Dağılım çubukları grafiği, kesikli nicel verilerde ve nitel verilerde kullanılır. Çubuk grafiğinde sınıflar, tabanları eşit ve birbirine bitişik olmayan dikdörtgenlerle temsil edilir. Tablo 1.4. de verilen frekans tablosu için çubuk grafiği aşağıda verilmiştir.



Grafik1.1. Hastaların hastalık türlerine göre grafiği

1.4.2. Çizgi Diyagramı

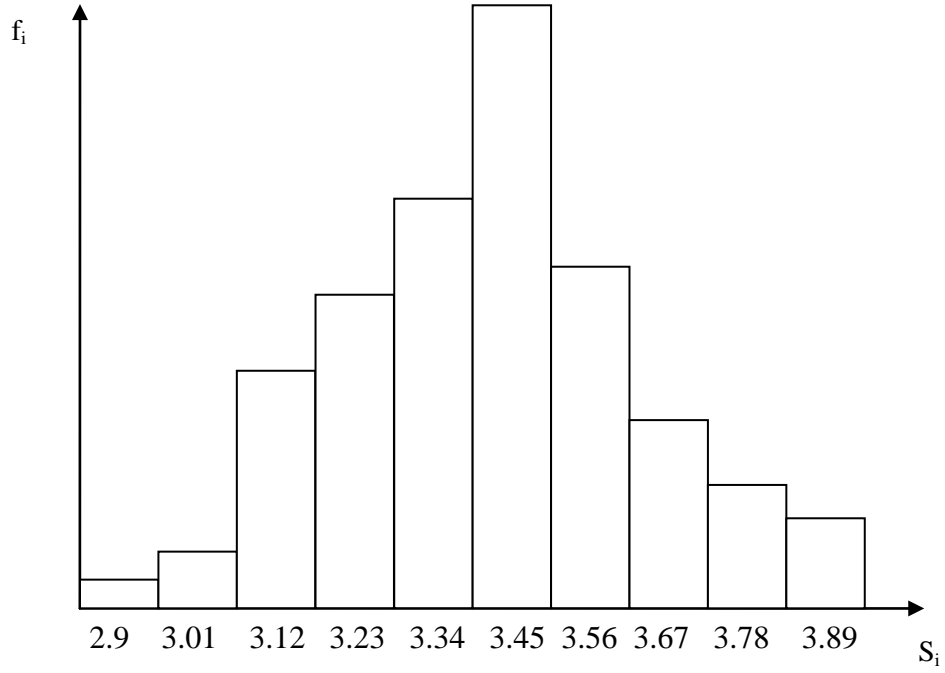
Frekans dağılımında her sınıfın düzlemde bir nokta ile temsil edilip, sonra bu noktaların birleştirilmesiyle elde edilen şekle çizgi diyagramı denir. Sürekli veriler için kullanılan bir grafik türüdür. Tablo 1.2. de verilen frekans tablosu için çizgi diyagramı aşağıda verilmiştir.



Grafik 1.2. 120 erkek bebeğin doğum ağırlıklarının grafiği

1.4.3 Histogram(Dağılım Dikdörtgenleri)

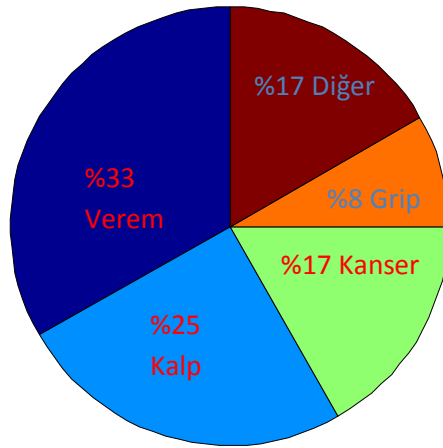
Histogram, koordinat sisteminde, tabanları frekans tablosundaki her bir sınıfın sınıf büyüklüğü, yükseklikleri bulunduğu sınıfın frekanslarıyla orantılı olarak yan yana çizilen dikdörtgenlerden oluşur. Tablo 1.2. de verilen frekans tablosu için histogram aşağıda verilmiştir.



Grafik 1.3. 120 erkek bebeğin doğum ağırlıklarının grafiği

1.4.4 Daire (Pasta)Grafığı

Her sınıfa düşen frekansın bir dairenin parçası ile gösterildiği grafik türüdür. Bu grafiği çizebilmek için görel frekanslar hesaplanır. Her sınıfa ilişkin görel frekans 360^0 ile çarpılarak o sınıfa ilişkin daire dilimleri bulunur. Tüm sınıflar için yapıldığında daire tamamlanmış olur. Daha çok sınıflandırılabilen verilerde kullanılır. Tablo 1.4. de verilen frekans tablosu için daire grafiği aşağıda verilmiştir.



Grafik 1.2. Hastaların hastalık türlerine göre dağılımı