Bölüm 4
Dağılış Ölçüleri

Doç.Dr. Suat ŞAHİNLER

Yer ölçüleri gözlemlerin sadece merkezini göstermekte olup, bu merkez etrafındaki dağılışın şekli hakkında bir fikir vermezler.

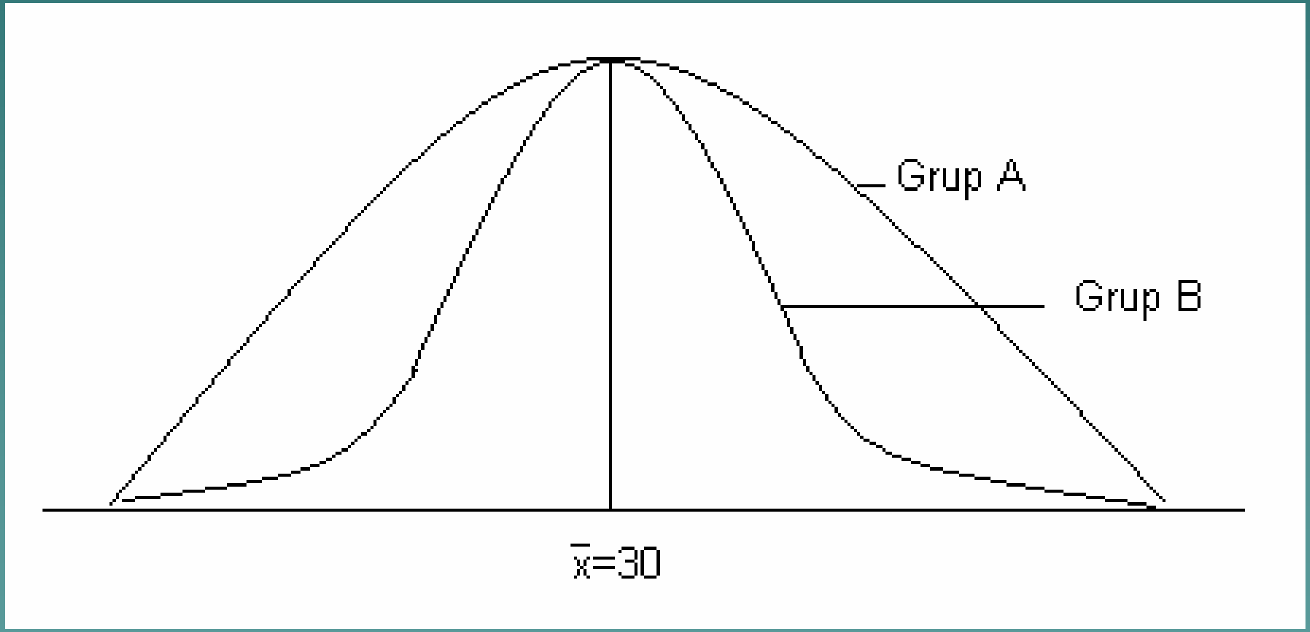
Örneğin aynı ortalamaya sahip iki grubun gösterdiği dağılışın şekli birbirinden farklı olabilir. Bu durumu daha iyi açıklayabilmek için aşağıdaki iki grup gözlemi ele alalım.

X_A={20,28,30,25,40,22,45}, $\bar{x}_A = 30$

X_B={27,28,32,30,33,26,34}, $\bar{x}_B = 30$

Aşağıdaki her iki grubun ortalaması aynı olduğu halde gözlemlerin gösterdiği dağılışın şekli birbirinden farklıdır.

Şekilden de görüldüğü üzere grup B'nin gözlem değeri grup A'ya göre daha fazla ortalama etrafında toplanmıştır. Başka bir deyişle grup A'daki gözlemler ortalamadan en fazla 45-30 = 15 birim, fakat grup B'de ise 34-30 = 4 birim farklılık göstermektedirler.



Buraya kadar yapılan açıklamalardan anlaşılacağı üzere, yer ölçüleri verilerin tüm özelliklerini tek başına yansıtabilecek yeterli bir ölçü değildirler. Bundan dolayı, verilerin diğer bazı özelliklerini yansıtabilecek başka ölçülerin kullanılmasına gerek vardır.

Bu bölümde istatistikte en çok kullanılan dağılış ölçülerinden;

- Değişim Genişliği
- Değişim Oranı
- Kantiller Arası Genişlik
- Ortalamadan Mutlak Sapma
- Varyans
- Standart Sapma
- Standart Hata
- Değişim (Varyasyon) Katsayısı
- Standart Değişken

üzerinde durulacaktır.

Değişim Genişliği

Değişim genişliği, gözlemlerin en büyük değeri (Xmax) ile en küçük değeri (Xmin) arasındaki farktır. Bu tanıma göre,

Değişim Genişliği (DG) = Xmax – Xmin

şeklinde hesaplanır.

Örnek:

$X_A=\{20,28,30,25,40,22,45\}$,

$X_B=\{27,28,32,30,33,26,34\}$,

Grupları için değişim genişliğini bulunuz.

$DG_A= 45-20 = 25$

$DG_B= 34-26 = 8$

Değişim genişliğinin hesaplanması basit ve anlaşılması kolay olması yanında değişim genişliğini verirken bunun hesaplanmasında kullanılan iki sayıyı da belirtmekte yarar vardır. Ancak bu durumda okuyucuya daha iyi fikir verir.

Değişim genişliği iki aşırı uç değere bağlı olarak hesaplanışı nedeniyle uç değerler arasındaki diğer gözlemlerin hangi bölgede yoğunlaştığı konusunda bir bilgi vermez.

Bunun yanında, değişim genişliği örnek büyüklüğü ile birlikte büyüyebileceğinden, eşit sayıda gözlem bulundurmayan örneklerin karşılaştırılması anlamlı bir sonuç vermez.

Son olarak değişim genişliği, matematik işlemlere de elverişli değildir

Varyans ve Standart Sapma

Varyans: gözlemlerin aritmetik ortalamadan olan sapmalarının kareler ortalaması şeklinde tanımlanabilir. Bu tanım çerçevesinde varyans formülü ;

Populasyon için

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$$

Örnek için

$$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

Burada;

σ^2 : Populasyon varyansını,

μ : Populasyon ortalamasını,

N : populasyon genişliğini,

göstermektedir.

s^2 : Örnek varyansını

$\bar{x}$: Örnek ortalamasını

n : Örnek genişliğini

İstatistikte dağılış ölçüsü olarak kullanılmaya en uygun ve en elverişli ölçü; varyans ve bunun karekökü olan standart sapmadır. Her iki ölçü bir populasyonu veya bir örneği oluşturan tüm gözlemler dikkate alınarak hesaplanır.

Genellikle varyansın yorumlamak oldukça güçtür. Çünkü varyansın birimi, verilerin ifade edildiği ölçü biriminin karesidir. Yani, veriler kg, gr, m, cm ile ifade edilmişse varyansın birimi sırasıyla kg2 , gr2 , m2 , cm2 olur . Bu da bir anlam taşımaz. Bundan dolayı varyansın karekökü alınır ve buna standart sapma denir. Standart sapmanın birimi verilerin ifade edildiği ölçü birimi ile aynıdır (kg, gr, m, cm). Standart sapma varyansın karekökü olduğuna göre;

Populasyon için

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

Örnek için

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

şeklinde hesaplanır.

Standart sapma: gözlemlerin aritmetik ortalamadan olan sapmalarının ortalaması şeklinde tanımlanabilir. Yukarıdaki formül kullanılarak varyansın hesaplaması özellikle gözlem sayısı fazla olduğunda güç olur. Bu nedenle varyansın

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2 / n}{n-1}$$

formülü ile hesaplanması daha kolaydır. Her iki formül de aynı sonucu verir.

Örnek yukarıda verilen Grup A ve B için varyansı ve standart sapmayı hesaplayınız.

X_A={20,28,30,25,40,22,45}, $\overline{x}_A = 30$

X_B={27,28,32,30,33,26,34}, $\overline{x}_B = 30$

Örnek: Kırk koyunun ağırlıkları ile ilgili frekans tablosu için aritmetik ortalamayı hesaplayınız.

A Grubu				B Grubu			
x_i	x_i^2	X_i-30	$(X_i-30)^2$	x_i	x_i^2	X_i-30	$(X_i-30)^2$
20	400	-10	100	27	729	-3	9
28	784	-2	4	28	784	-2	4
30	900	0	0	32	1024	2	4
25	625	-5	25	30	900	0	0
40	1600	10	100	33	1089	3	9
22	484	-8	64	26	676	-4	16
45	2025	15	225	34	1156	4	16
210	6818	0	518	210	6358	0	58

A Grubu için yukarıdaki 1. formülde bilinenler yerine yazılırsa varyans;

$$S_A^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{518}{7-1} = \frac{518}{6} = 86.33$$

veya 2. formülde bilinenler yerine yazılırsa varyans

$$s_A^2 = \frac{\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2 / n}{n-1} = \frac{6818 - 210^2 / 7}{7-1} = \frac{6818 - 6300}{6} = 86.33$$

olarak hesaplanır.

B Grubu için 1. formülde bilinenler yerine yazılırsa varyans;

$$S_B^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{58}{7-1} = \frac{58}{6} = 9.67$$

veya 2. formülde bilinenler yerine yazılırsa varyans

$$s_B^2 = \frac{\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2 / n}{n-1} = \frac{6358 - 210^2 / 7}{7-1} = \frac{6358 - 6300}{6} = 9.67$$

olarak hesaplanır. (Bu değerleri yorumlayınız)

Her iki grup için standart sapma değerleri ise

$$S_A = \sqrt{S_A^2} = \sqrt{86.33} = 9.29$$

$$S_B = \sqrt{S_B^2} = \sqrt{9.67} = 3.11$$

olarak hesaplanırlar. (Bu değerleri yorumlayınız)

Sınıflandırılmış verilerde varyansın hesaplanması;
Sınıflandırılmış verilerde varyans;

$$S^2 = \frac{\sum_{j=1}^{k} f_j (x_j - \bar{x})^2}{\sum_{j=1}^{k} f_j - 1} = \frac{\sum_{j=1}^{k} f_j x_j^2 - (\sum_{j=1}^{k} f_j x_j)^2 / \sum_{j=1}^{k} f_j}{\sum_{j=1}^{k} f_j - 1} = \frac{\sum fx^2 - (\sum fx)^2 / \sum f}{\sum f - 1}$$

formülü ile hesaplanır.

Örnek: 40 koyunun ağırlığı ile ilgili frekans tablosu için varyansı hesaplayınız.

Önce sınıf değerleri bulunur ve formül gereği frekanslarla çarpılır ve toplanırsa ve sınıf değerlerinin karelerinin frekanslarla çarpılırsa…

Sınıf Limitleri (Koy. Ağr.(kg))		f_j (Koy.Say)	Sınıf Değeri (x_j)	f_jx_j	$f_jx_j^2$
39	42	1	40.5	40.5	1640.25
43	46	4	44.5	178.0	7921.00
47	50	8	48.5	388.0	18818.00
51	54	11	52.5	577.5	30318.75
55	58	7	56.5	395.5	22345.75
59	62	5	60.5	302.5	18301.25
63	66	2	64.5	129	8320.50
67	70	2	68.5	137	9384.50
Σ		40	-	2148	117050.00

Sınıflandırılmış verilerde varyans formülündeki tabloda hesaplanan unsurlar yerine yazılırsa varyans;

$$S^2 = \frac{\sum_{j=1}^{k} f_j x_j^2 - (\sum_{j=1}^{k} f_j x_j)^2 / \sum_{j=1}^{k} f_j}{\sum_{j=1}^{k} f_j - 1} = \frac{117050 - 2148^2 / 40}{40 - 1} = \frac{117050 - 115347.6}{39} = \frac{1702.4}{39} = 43.65$$

olarak, Standart sapması ise;

$$S = \sqrt{S^2} = \sqrt{43.65} = 6.61\,kg$$

olarak hesaplanır.

Standart Hata

Ortalamanın standart hatası olarak da bilinir ve

$$S_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n}}$$

formülü ile hesaplanır.

Standart hata, gözlem değerleri hangi ölçü birimi ile ölçülmüş ise o ölçü birimi ile ifade edilir.

Örnek: 40 koyunun ağırlığına ait standart hata değerini hesaplayınız.

$$S_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{43.65}{40}} = \frac{6.61}{\sqrt{40}} = 1.045$$

Varyasyon Katsayısı (Değişim Katsayısı)

Standart sapma (veya varyans) bir değişim ölçüsü olarak iki gruba ait gözlemlere bakarak hangi grubun daha homojen olduğunun belirlenmesinde her zaman yeterli bir ölçü olmayabilir. Çünkü bazen yapılan ölçümlerin büyüklüğü standart sapmayı ortalamadan daha fazla etkileyebilir. Diğer bir ifadeyle

- gözlemlerin büyüklüğü standart sapmayı etkilediği biliniyor veya öyle düşünülüyorsa,
- gözlemler farklı ölçü birimi ile ifade edilmişlerse

standart sapma yeterli bir değişim ölçüsü olmayıp yanıltıcı sonuçlara neden olabilir.

İşte bu gibi durumlarda, yani gözlem değerlerinin büyüklüğünden ileri gelen farklılığı ortadan kaldırmak hem de farklı ölçü birimi ile ifade edilmiş gözlem değerlerini karşılaştırılabilir duruma getirmek için yeni bir değişim ölçüsü kullanılması gerekmektedir. Bu da varyasyon katsayısı (değişim katsayısı) olup bu, standart sapmanın ortalamaya oranının yüzle çarpılmasıyla bulunur ve yüzde ile ifade olunur. Yani;

Varyasyon katsayısı

$$VK = \frac{S}{\bar{x}} * 100$$

formülü ile hesaplanır.

Varyasyon katsayısı ne kadar küçük olursa denemenin veya çalışmanın sonucuna olan güvenilirlik o oranda artar. Yalnız biyolojik çalışmalarda bu değerin % 30 ‘un altında olması istenir.
Aynı konuda yapılan çalışmalardan hangisinin sonucuna daha çok güvenmemiz hususunda karar vermede yardımcı olan tek kriter varyasyon katsayısıdır. Düşük varyasyon katsayılı çalışma, diğerlerine nispetle daha sağlıklı yürütülmüş demektir.
Son olarak düşük varyasyon katsayılı grup, grubu oluşturan birey veya birimler arasındaki varyasyonun az olduğunu yani homojen bir grup olduğunu ortaya koyar. Aksi halde ise heterojen bir gruptur.

Örnek: Koyun ve sığırların ağırlıkları ile ilgili standart sapmalar sırasıyla 6.4 ve 58 kg ve ortalamalar ise sırasıyla 56 ve 571 kg olarak bulunmuştur. Bu duruma göre varyasyonun hangi tür için daha fazla olduğu söylenebilir.

$$VK_K = \frac{S}{\bar{x}} * 100 = \frac{6.4}{56} * 100 = \%11.42$$

$$VK_S = \frac{S}{\bar{x}} * 100 = \frac{58}{571} * 100 = \%10.16$$

Koyunlar için varyasyon katsayısı %11.42, sığırlar için ise %10.16 olarak hesaplanmıştır. Bu durumda, standart sapma değerlerine göre ağırlık bakımından koyunlar daha homojen görünürken, Varyasyon katsayılarına göre ağırlık bakımından sığırlar daha homojendir, yani varyasyon (değişim) koyunlarda, sığırlara göre daha fazladır.

Standart Değişken (Standart Puan)

Gözlem değerlerinin ortalamadan olan farklarının standart sapma cinsinden ifadesidir şeklinde tanımlanabilir. Standart değişken, gözlemlerin ölçüldüğü ölçü biriminden bağımsız olup birimi yoktur. Yalnız standart sapma cinsinden ifade olunup;

$$Z = \frac{x_i - \mu}{\sigma}$$

veya

$$Z = \frac{x_i - \bar{x}}{S}$$

formülü ile hesaplanır.

Örnek .4.12. İstatistik dersi alan 364 öğrencinin sınavda aldıkları notların ortalaması 70 puan ve standart sapması 10 puan olarak bulunmuştur. Bu sınavda A ve B öğrencileri sırasıyla 80 ve 50 puan almışlardır. Buna göre her iki öğrencinin standart puanlarını bulunuz.

$$Z_A = \frac{x_A - \mu}{\sigma} = \frac{80-70}{10} = 1$$

$$Z_B = \frac{x_B - \mu}{\sigma} = \frac{50-70}{10} = -2$$

puan olarak hesaplanır. Diğer bir ifadeyle, A öğrencisi ortalamadan bir standart puan yüksek, fakat B öğrencisi ise iki standart puan düşük puan almıştır.