Sosyal Bilimlerde İstatistik

Bölüm III: Frekans Dağılımları, Merkezi Eğilim ve Yayılım Ölçüleri

Nihan Acar Denizli[1]

[1]Mimar Sinan Güzel Sanatlar Üniversitesi

İÇERİK

Frekans Dağılımları

Herhangi bir değişkenin aldığı değerin anakütle ya da örneklemde ne sıklıkla gözlendiğini ifade eden sayıya *frekans* denir.

Sınıflama ya da sıralama ölçme düzeyinde ölçülmüş (nitel) değişkenlerin her bir sınıfındaki tekrar sayısının gösterildiği tablolara ise *frekans tabloları* adı verilir.

Örnek:

Eğitim Durumu	Frekans(Sıklık)
İlköğretim	15
Lise	55
Yüksekokul	5
Üniversite	35
Lisansüstü	10
Toplam	120

Frekans Dağılımları

Nicel Bir Değişkenin Sınıflandırılması: Eşit aralıklı ya da oran ölçüm düzeyinde ölçülmüş değerler sınıflandırılarak frekans tabloları oluşturulabilir. Bunun için,

- Değer aralığı uzunluğu: Örneklemdeki en büyük değer ile en küçük değer arasındaki fark.
- Sınıf sayısı (k): Araştırmacı tarafından oluşturulmak istenen sınıf sayısı.

bilinmelidir.

Buradan, istenilen sınıf sayısı için kullanılması gereken sınıf aralığı (c)

$$c = \frac{\text{değer aralığı uzunluğu}}{\text{sınıf sayısı}}$$

olur.

Nicel Bir Değişkenin Sınıflandırılması Örneği

Örnek: 15 farklı öğrencinin bir ayda okudukları kitap sayıları şu şekildedir:

5 1 2 3 3 6 1 4 7 5 3 4 2 2 6

Okunan Kitap Sayısı	Frekans	Birikimli Frekans
1	2	2
2	3	5
3	3	8
4	2	10
5	2	12
6	2	14
7	1	15

Nicel Bir Değişkenin Sınıflandırılması Örneği

Örnek: 15 farklı öğrencinin bir ayda okudukları kitap sayıları şu şekildedir:

5 1 2 3 3 6 1 4 7 5 3 4 2 2 6

Bu değerleri 3 sınıflı bir frekans tablosunda göstermek isteyelim:

$$\text{değişim aralığı uzunluğu} = 7 - 1 = 6$$

$$c = 6/3 = 2$$

Okunan Kitap Sayısı	Frekans
1-3 arası kitap	8
4-6 arası kitap	6
7 ve daha fazla	1

Sınıf sayısı $k = 2$ alınırsa, c=3 olur.

Okunan Kitap Sayısı	Frekans
1-4 arası kitap	10
5 ve daha fazla kitap	5

İÇERİK

Merkezi Eğilim ve Yayılım Ölçüleri

Merkezi Eğilim Ölçüleri: Değerlerin etrafında kümelendiği, dağılımın ortalamasına dair bilgi veren ölçüler (Ortalama, medyan, mod).

Merkezi Yayılım Ölçüleri: Değerlerin merkezin / ortalamanın etrafında nasıl dağıldığına dair bilgi veren ölçüler (Değişim aralığı, standart sapma, varyans).

Merkezi Eğilim Ölçüleri: Ortalama

Aritmetik Ortalama: Alınan tüm değerlerin toplanıp, toplam birim sayısına bölünmesiyle hesaplanan ortalama türü. $x_1, x_2, ..., x_n$, n gözlemden oluşan veri seti için aritmetik ortalama:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Sınıflandırılmış veriler için aritmetik ortalama grup tekrar sayısının (f_i) ilgili grup değeri (x_i) ile çarpımı ya da sınıf ortasının frekans değeri ile çarpımı üzerinden hesaplanır.

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_i f_i}$$

Merkezi Eğilim Ölçüleri: Ortalama

Örnek: 15 öğrencinin bir ayda okudukları kiap sayılarının aritmetik ortalamasının hesabı.

$$\bar{x} = \frac{\sum_{i=1}^{15} x_i}{15} = \frac{5+1+2+3+\ldots+6}{15} = 3.6$$

3 sınıfa ayrılmış veri seti için aritmetik ortalama sınıf ortasındaki değer ile bu sınıfa ait frekansın çarpımının toplam birim sayısına bölünmesiyle hesaplanır:

$$\bar{x} = \frac{(2*8)+(5*6)+(7*1)}{15} = 3.53$$

Aritmetik Ortalamanın Özellikleri

- Aritmetik ortalamadan farkların toplamı sıfırdır.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0.$$

- Ortalamadan sapmanın toplamı minimumdur.
- Aşırı / uç değerlerden etkilenir.
- Hesabında tüm birimler kullanıldığından açık sınıflı serilerde hesaplanamaz.
- Sadece nicel (eşit aralıklı ve oran ölçeğinde ölçülmüş) değişkenler için kullanılır.

Merkezi Eğilim Ölçüleri: Ağırlıklı Ortalama

Ağırlıklı (Tartılı) Ortalama: Veri grubundaki değerlerin farklı ağırlıklara ya da öneme sahip olduğu durumda kullanılır. Aritmetik ortalamaın özel bir durumudur. $x_1, x_2, ..., x_n$, n gözlemden oluşan veri seti için ağırlıklı

ortalama x_i gözlem değerini ve w_i ağırlıkları ifade etmek üzere şu şekilde hesaplanır:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i w_i}{w_i}$$

Örnek: Öğrencilerin genel akademik not ortalamalarının hesaplanması.

Merkezi Eğilim Ölçüleri: Geometrik Ortalama

Geometrik Ortalama: Oransal farkların önemli olduğu durumlarda veya artma ya da azalma eğiliminde olan serilerin ortalamasının hesabında kullanılır. Gözlenen değerlerin çarpımının gözlem sayısı derecesinden kökünün alınmasıyla hesaplanır. 0 ve/veya negatif değerli birimle için geometrik ortalama hesaplanamaz. $x_1, x_2, ..., x_n$, n gözlemden oluşan veri seti için geometrik ortalama:

$$\bar{x}_{GO} = \sqrt[n]{x_1 * x_2 * \ldots * x_n} = \sqrt[n]{\prod_{i}^{n} x_i}$$

$$\log \bar{x}_{GO} = \frac{1}{n} \sum_{i=1}^{n} \log x_i$$

Örnek: Fiyat artış oranlarının, nüfus artış oranlarının ortalaması.

Merkezi Eğilim Ölçüleri: Geometrik Ortalama

Örnek: İstanbul için verilen yıllara göre nüfus artış oranlarına göre 2008-2011 yılları arasındaki ortalama nüfus artışını hesaplayınız.

Yıl	İstanbul Nüfusu	Artış Hızı
2008	12.697.164	% 0.98
2009	12.915.158	% 1.72
2010	13.255.685	% 2.64
2011	13.624.240	% 2.78
2012	13.854.740	% 1.69
2013	14.160.467	% 2.21
2014	14.377.018	% 1.53
2015	14.657.434	% 1.95
2016	14.804.116	% 1.00

Merkezi Eğilim Ölçüleri: Geometrik Ortalama

Çözüm:

$$\begin{aligned}\log \bar{x}_{GO} &= \frac{1}{4}\sum_{i=1}^{4} \log x_i \\ &= \frac{1}{4}\sum [\log (0.98) + \log (1.72) + \log (2.64) + \log (2.78)] \\ &= \frac{1}{4}(1.092) = 0.273\end{aligned}$$

$$\bar{x}_{GO} = 10^{0.273} = 1.87$$

Merkezi Eğilim Ölçüleri: Ortanca (Medyan)

Sıralandırılmış verilerin ortasındaki değer.

Veri setinde tek sayıda gözlem varsa, $\frac{n+1}{2}$. değer,
çift sayıda gözlem varsa ortada kalan iki değerin ($\frac{n}{2}$. ve $\frac{n+2}{2}$. değerlerin) ortalaması olarak hesaplanır.

Sınıflandırılmış veriler için tek sayıda gözlem varsa f frekansı ifade etmek üzere $\frac{f+1}{2}$. gözlemin, çift sayıda gözlem varsa $\frac{f}{2}$. ve $\frac{f+2}{2}$. gözlemlerin bulunduğu sınıftır.

Alt ve üst sınırlar gözönüne alınarak oluşturulmuş frekans tablolarından ortanca (medyan) formül aracılıgıyla hesaplanır:

$$\text{Medyan} = L + \frac{c_i}{f_i}\left(\frac{n}{2} - d\right)$$

L: medyan sınıf alt değeri, c: Medyan sınıfı aralığı, f: Medyan sınıfı frekansı, d: Medyan sınıfından önceki sınıfların frekansları toplamı

Ortancanın (Medyanın) Özellikleri

- Hesaplanması kolaydır.
- Aşırı / uç değerlerden etkilenmez.
- Açık sınıflı ya da sınıf aralıkları farklı olan serilerde de kullanılabilir.
- Gözlemleri dengeleyen bir ölçüdür.
- Eşit aralıklı ve oran ölçeğinin yanısıra sıralama (ordinal) ölçeğinde ölçülmüş değişkenler için de hesaplanabilir.

Merkezi Eğilim Ölçüleri: Tepe Değer (Mod)

Veri setinde en çok tekrarlanan gözlem değerine eşittir.

Sınıflandırılmış veriler için en yüksek frekans değerine sahip olan kategori *mod sınıfı* olarak adlandırılır.

Sınıfların aralık biçiminde tanımlandığı frekans dağılımlarında formül aracılığıyla hesaplanır.

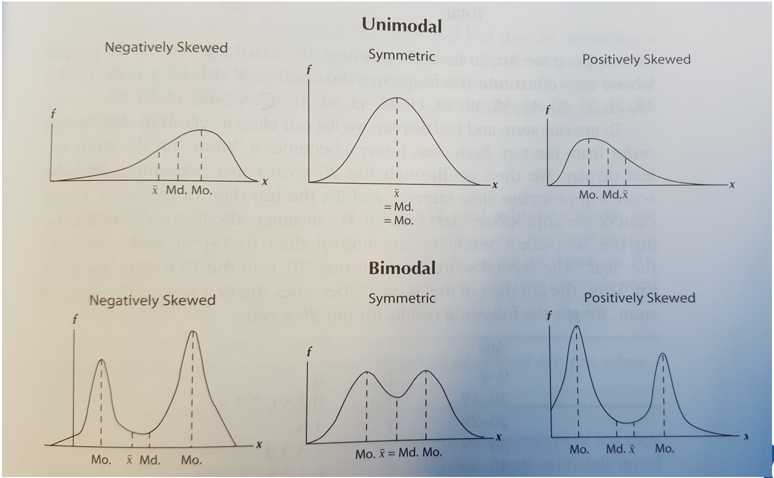
$$\text{Mod} = L + c\frac{f_\text{s}}{f_\text{s} + f_\text{ö}}$$

L: mod sınıfı alt değeri, c: Mod sınıfı aralığı, f_s: Mod sınıfından sonraki sınıfın frekansı, $f_ö$: Mod sınıfından önceki sınıfın frekansı

Tepe Değerin (Modun) Özellikleri

- Hesaplanması kolaydır.
- Açık sınıflı gruplanmış verilerde de hesaplanabilir.
- Hem nicel hem nitel veriler için hesaplanabilir. Tüm ölçüm düzeylerinde (sınıflama, sıralama, eşit aralıklı ve oran) ölçülmüş veriler için hesaplanabilir.

Simetrik - Çarpık, Tek Modlu - Çift Modlu Dağılımlar



Merkezi Yayılım Ölçüleri

Değişim Aralığı (Range): Veri setindeki en büyük değer ile en küçük değerin arasındaki fark olarak hesaplanan en basit yayılım ölçüsüdür.

Ortalamadan Sapma (Mean Deviation): Bir değerin ortalamadan sapmasının ortalama uzaklığını verir. Veri setindeki değerlerin ortalamadan farklarının mutlak değerinin ortalaması olarak hesaplanır.

$$\text{O.S.} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Merkezi Yayılım Ölçüleri

Varyans (Variance): Aritmetik ortalama ile gözlem değerleri arasındaki farkların kareleri toplamına dayanan ve en sık kullanılan yayılım ölçüsüdür.Gözlem değeri aritmetik ortalamaya ne kadar yakın ise varyans o kadar küçük olur. Gözlem değerlerinin ortalamada etrafında ne derece yaygın olduğunu gösterir. Birimi orjinal değişken biriminin karesidir. Anakütle için σ^2, örneklem için s^2 olarak gösterilir ve şu şekilde hesaplanır:

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

Değerler geniş bir aralık içinde dağılıyorsa büyük, dar bir aralık içinde dağılıyorsa küçük değer alır.

Merkezi Yayılım Ölçüleri

Standart Sapma (Standart Deviation): Verilerin ortalamaya göre yayılımını gösteren bir ölçüdür. Varyansın pozitif kare kökü olarak hesaplanır. Eğer verilerin çoğu ortalamaya yakın değerler almış ise standart sapma değeri küçük, birçok veri ortalamadan uzak değerler almışsa standart sapma değeri büyüktür. Bütün gözlem değerleri aynı değeri almış ise standart sapma sıfıra eşittir.

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

Merkezi Yayılım Ölçüleri

Değişim Katsayısı: Farklı birimlerdeki değişkenlerin ya da aynı değişkene ait ortalamaları farklı iki ya da daha fazla grubun homojenliğini (değişkenliğini) karşılaştırmak için *değişim katsayısı* kullanılır.

$$\text{D.K.} = \frac{s}{\bar{x}}$$

Çarpıklık Ölçüsü (Skewness): Gözlem değerlerinin dağılımın hangi aralığında daha çok gözlendiğine ilişkin bilgi verir. Gözlem değerlerinin oralama etrafındaki dağılımı ne kadar simetrik olursa çarpıklık değeri sıfıra o kadar yakın olur.

Basıklık Ölçüsü (Kurtosis): Gözlem değerlerinin oluşturduğu dağılımın normal dağılıma göre daha sivri ya da basık olup olmadığı konusunda bilgi verir. Pozitif değer alıyor ise dağılım sivri, negatif değer alıyor ise dağılım basıktır.

Örnekler

Bir öğrencinin son 10 aydaki aylık gideri:

Kişi	Not	$(x_i - \bar{x})^2$
1	500	$(-200)^2$
2	650	$(-50)^2$
3	700	0
4	800	$(100)^2$
5	550	$(-150)^2$
6	850	$(150)^2$
7	750	$(50)^2$
8	700	0
9	900	$(200)^2$
10	600	$(-100)^2$
Toplam	7000	150000

Örnekler

Bir öğrencinin son 10 aydaki aylık gideri:

Varyans:

$$s^2 = \frac{150000}{9} = 16666.66$$

Standart sapma:

$$\sqrt{16666.66} = 129.09$$

Öğrencinin aylık ortalama $700(s.d. = 129.09)$ TL gideri vardır.

Örnek: Değişim Katsayısı

Örnek: Aylık ortalama gideri 700 TL, standart sapması 140 TL olan öğrenci ile aylık ortalama gideri 600 Euro standart sapması 150 Euro olan öğrenciden hangisinin aylık gideri daha homojendir?

$$\text{D.K.1} = \frac{s_1}{\bar{x}_1} = 140/700 = 0.2$$

$$\text{D.K.2} = \frac{s_2}{\bar{x}_2} = 150/600 = 0.25$$

Aylık harcaması daha fazla olan öaylık gideri daha homojendir.

KAYNAKLAR

- Nuran Bayram (2009). Sosyal Bilimlerde SPSS ile Veri Analizi, Ezgi Kitabevi.
- Abdullah Can (2014). SPSS ile Bilimsel Araştırma Sürecinde Nicel Veri Analizi, Pegem Akademi Yayınları, 2. Baskı.
- Kazım Özdamar (2013). Paket Programlar İle İstatistiksel Veri Analizi, Nisan Kitabevi.
- Neyran Orhunbilge (2014). Tanımsal İstatistik: Olasılık ve Olasılık Dağılımları, 2.Baskı, Nobel Yayınevi.
- Özkan Ünver, Hamza Gamgam ve Bülent Altunkaynak (2006). Temel İstatistik Yöntemler Hipotez Testleri-İlişki Katsayıları- Regresyon Analizi, 7.Baskı, Seçkin Yayıncılık.
- R. Mark Sirkin (2005). Statistics for the Social Sciences, SAGE.
- Sait Gürbüz ve Faruk Şahin (2016). Sosyal Bilimlerde Araştırma Yöntemleri Felsefe-Yöntem-Analiz, 3.Baskı, Seçkin Yayıncılık.
- Ümit Şenesen (2013). İstatistik: Sayıların Arkasını Anlamak, Dördüncü Basım, Literatür Yayıncılık Dağıtım.
- Yahşi Yazıcıoğlu ve Samiye Erdoğan (2014). SPSS Uygulamalı Bilimsel Araştırma Yöntemleri, Detay Yayıncılık.