TEMEL İSTATİSTİK

Merkezi Eğilim Ölçüleri

Prof. Dr. Ezel Tavşancıl

Geçen haftadan hatırlatma…

- Geçen hafta, betimsel istatistiklerden bir olan dağılımın betimlenmesi ve grafik ya da tablo ile gösterilmesi anlatıldı.
- Bu hafta betimsel istatistikler başlığı altındaki diğer iki konudan biri: **merkezî eğilim ölçüleri** işlenecek.

MERKEZİ EĞİLİM ÖLÇÜLERİ…

- DİĞER İSİMLENDİRMELER:
  - Vasat Ölçüleri (Tan, 2016)
  - Konum Ölçüsü (Howitt & Cramer, 1997)
  - Merkeze Yığılma (Baykul, 1999)

- Grup ölçümüne ilişkin tipik değer
- İlgilen değişkene ilişkin bir grup ölçümün ortalama durumunu yansıtır

(Howitt & Cramer, 1997)

- ÖRN: Alınan istatistik testi sonuçlarına göre ortalama bir öğrenci düşünülerek tek bir ölçü ile betimlenebilir. İlgilenilen değişkene ilişkin ölçek üzerinde bir değer veya noktaya karşılıktır (Büyüköztürk, Çokluk ve Köklü, 2018).

…MERKEZİ EĞİLİM ÖLÇÜLERİ

- En sık kullanılan üç merkezî eğilim ölçüsü:
  - Aritmetik Ortalama
  - Ortanca (Medyan)
  - Mod (Tepe Değer)

- Hangisinin kullanılacağının belirlenmesi için ilgili ölçülerin temel nitelikleri ile kullanıldıkları durumların, güçlü ve zayıf yönlerinin ve de nasıl hesaplandıklarının bilinmesi gerekir.

(Tan, 2016)

Mod (Tepe Değer)…

- Bir değişkenle ilgili ölçümlerden en çok tekrar edilen ölçme sonucuna denir (Büyüköztürk ve diğerleri, 2018).
- Frekansı en büyük olan puana denir.

- **! DİKKAT**: Değerler, aynı değişkene ilişkin ölçümler dizisi içinde değerlendirilmeli

- Mod, sınıflama ölçeği düzeyindeki veriler için uygun bir merkezî eğilim ölçüsüdür (Tan, 2016).

- **ÖRN: Veri Seti**: *60, 72, 82, 72, 61, 81, 72* > **Mod**: *72*’dir.

- Bazı durumlarda, dağılımın iki veya daha fazla modu olabilir. Bu durumda dağılıma iki modlu ya da çok modlu modlu denir (Tan, 2016). Özellikle dağılımın normalden uzaklaştığı durumlarda (Büyüköztürk ve diğerleri, 2018).

...Mod (Tepe Değer)...

- Gruplandırılmış verilerde mod, frekansın en yüksek olduğu puan aralığının orta noktasıdır.

Mod, 54-62 puan aralığının orta noktasıdır: 58

Tablo 3.2: Öğrencilerin Türkçe Başarı Testi Puanları

Puan Aralığı	f	X_0	fX_0	X_0^2	fX_0^2	x'	fx'	$(x')^2$	$f(x')^2$	t_f
90-98	2	94	188	8836	17672	9	18	81	162	100
81-89	6	85	510	7225	43350	8	48	64	384	98
72-80	9	76	684	5776	51984	7	63	49	441	92
63-71	12	67	804	4489	53868	6	72	36	432	83
54-62	17	58	986	3364	57188	5	85	25	425	71
45-53	13	49	637	2401	31213	4	52	16	208	54
36-44	15	40	600	1600	24000	3	45	9	135	41
27-35	14	31	434	961	13454	2	28	4	56	26
18-26	8	22	176	484	3872	1	8	1	8	12
09-17	4	13	52	169	676	0	0	0	0	4
Σ =			5071		297277		419		2251	

(Büyüköztürk ve diğerleri, 2018).

...Mod (Tepe Değer)...

Bir seride en çok tekrarlanan değere “Mod” denir.

Örnek: 10 öğrencinin ağırlıklarından oluşan seride mod;
72 80 58 60 65 75 51 59 60 60
Mod:60 kg'dır. 60 değeri en fazla tekrarlanandır.
Görüldüğü gibi 3 tane 60 vardır. Bu tür serilere tek modlu seri denir.

Örnek: 3 8 15 20 12 15 12 9 17
$Mo_1 = 12$
$Mo_2 = 15$
Görüldüğü gibi bu seride 2 tane 12 ve 2 tane de 15 değeri vardır.
Bunlardan birine birinci mod, diğerine ise ikinci mod değeri denir.
Bu tür serilere ise çok modlu seri denir.

…Mod (Tepe Değer)

AVANTAJLARI

1. Hesaplanması ve anlaşılması kolaydır.
2. Dağılımdaki aşırı değerlerden etkilenmez.

SAKINCALARI

1. Bazı dağılışlarda tepe değeri bulunmayabilir, bazılarında da birden fazla tepe değeri bulunabilir. İki tepe değeri bulunan dağılışlara **bimodal** dağılış adı verilir.
2. Tepe değeri aritmetik işlemler için elverişli değildir.

Aritmetik Ortalama

- Sadece ortalama dendiğinde, aritmetik ortalama anlaşılmalıdır (Arıcı, 1998).
- Gözlenen değerlerin tümü toplanarak gözlem sayısına bölündüğünde elde edilen değerdir.
- Evren için μ; örneklem için $\bar{X}$ ile gösterilir.

$$\bar{X} = \frac{\sum X}{n}$$

$$\bar{X} = \frac{X_1 + X_2 + X_3}{n}$$

BASİT GÖSTERİM

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{N} = \frac{X_1 + X_2 + X_3 + \ldots\ldots X_n}{N}$$

$\bar{X}$ = Aritmatik Ortalama

N = Toplam Veri Sayısı

X = Veri Değeri

DETAYLI GÖSTERİM

Tekrarlı Ölçümler için Aritmetik Ortalama

- Tek tek toplamak yerine gözlem değerlerini frekansları ile çarpıp toplayarak aritmetik ortalama hesaplayabiliriz.

$$\overline{X} = \frac{\sum fX}{n}$$

! **T**oplama işareti gereği, her bir gözlem değerini frekansı ile çarpıp çarpım sonuçlarını toplamalıyız.

- *Temel Prensip*: 5+5+5+5 =4*5 eşitliğidir.
- BİR KEZ GÖZLENEN DEĞER İÇİN f*X DEĞERİ KAÇA EŞİT OLACAKTIR?

ÖRNEK 1

Tablo 3.1: Öğrencilerin Okuma Hızı

Ölçüm (X)	Frekans (f)	fX
7	2	14
6	3	18
5	5	25
4	7	28
3	4	12
2	3	6
1	1	1
	n=25	$\sum fx = 104$

Buna göre öğrencilerin verilen metni ortalama okuma hızı,

$\overline{X} = \frac{\sum fX}{n} = \frac{104}{25}$ =4,16 olarak hesaplanır.

Gruplandırılmış Veriler için Aritmetik Ortalama

YÖNTEM I

- Her aralığın orta noktası (X_O) ve ham puanlar dikkate alınır.

$$\overline{X} = \frac{\sum fX_0}{n}$$

YÖNTEM II

- Grup aralıkları için kodlama yapılır ve yapay bir değer üretilir (X^{ι}). En küçük puan aralığı için 0'dan başlayarak sayı ataması yapılır. Sonraki aralıklar için sırasıyla 1, 2, 3 olarak atama devam eder.
- *T.O. Değeri*: İlk aralığın orta noktası $\left(\frac{Aralığın\ son\ değeri + Aralığın\ ilk\ değeri}{2}\right)$
- a: Aralık katsayısı

$$\overline{X} = T.O + \left(\frac{\sum fX^{\prime}}{n}\right)a$$

Tablo 3.2: Öğrencilerin Türkçe Başarı Testi Puanları

Puan Aralığı	f	X_0	fX_0	X_0^2	fX_0^2	x'	fx'	$(x')^2$	$f(x')^2$	t_f
90-98	2	94	188	8836	17672	9	18	81	162	100
81-89	6	85	510	7225	43350	8	48	64	384	98
72-80	9	76	684	5776	51984	7	63	49	441	92
63-71	12	67	804	4489	53868	6	72	36	432	83
54-62	17	58	986	3364	57188	5	85	25	425	71
45-53	13	49	637	2401	31213	4	52	16	208	54
36-44	15	40	600	1600	24000	3	45	9	135	41
27-35	14	31	434	961	13454	2	28	4	56	26
18-26	8	22	176	484	3872	1	8	1	8	12
09-17	4	13	52	169	676	0	0	0	0	4
Σ =			**5071**		**297277**		**419**		**2251**	

$\sum fX_0$ =5071 ve n=100 değerlerini Formül 3.3'te yerine koyduğumuzda:

$$\overline{X} = \frac{\sum fX_0}{n} = \frac{5071}{100} = 50.71 \text{ ve}$$

T.O=13, $\sum fx'$ =419, n=100 ve a=9 değerlerini Formül 3.4'te yerine koyduğumuzda:

$$\overline{X} = \text{T.O.} + \left(\frac{\sum fx'}{n}\right) a = 13 + \left(\frac{419}{100}\right)9 = 13 + (37.71) = 50.71 \text{ bulunur.}$$

Aritmetik Ortalamanın Özellikleri

AVANTAJLARI

1. Hesaplanması ve anlaşılması kolaydır.
2. Her dağılımda bir tane aritmetik ortalama vardır.
3. Aritmetik işlemler için elverişlidir.

SAKINCALARI

1. Dağılımdaki aşırı değerlerden ileri derecede etkilenir.
2. Dağılımdaki aşırı değerler aritmetik ortalamayı kendilerine doğru kaydırırlar.
3. Bu etkilenme aşırı değerlerin aşırılık ölçüsü ile doğru, dağılımdaki veri sayısıyla ters orantılıdır.
4. Ters yöndeki aşırı değerler birbirlerinin etkisini azaltır.

Ortanca…

- Küçükten büyüğe doğru sıralanmış bir ölçüm grubunun orta puanını gösterir.

- Ortanca verilerin dağılımının normalden uzak olması, sağa ya da sola çarpık olması durumunda kullanılır.

- Çünkü böyle durumlarda ortalama uç değerlerden etkilenirken, ortanca etkilenmez.

…Ortanca…

Önce veriler büyüklük sırasına dizilir.

i. Veri sayısı tek ise, n+1/2 sıra numaralı değer ortanca olarak alınır.

15 18 **21** 24 28

ii. Veri sayısı çift ise n/2 sıra numaralı değer ile bir sonraki değerin aritmetik ortalaması ortanca olarak kabul edilir.

15 18 **21** **24** 28 32

(21+24) / 2 = 22.5

Gruplandırılmış ve gruplandırılmamış, ancak tekrarlı ölçümler içeren veri seti için

$$\boxed{Ortanca = L + \left(\frac{\frac{n}{2} - t_{fa}}{f_b} \right) a}$$

- L: n/2 frekansın rast geldiği aralığın gerçek alt sınırı
- t_{fa}: n/2 frekansının rast geldiği aralığa kadar olan toplamalı frekans
- f_b: Bu aralığa karşılık gelen frekans

ÖRNEK 1

$n/2=25/2=12.5$
12.5. gözlemin denk geldiği aralık için:
f_b : Frekans sayısı=7
t_{fa} : Öncesinin toplamalı frekansı= 8

Tablo 3.3: Öğrencilerin Okuma Hızı

Ölçüm (X)	Frekans (f)	fX	Toplamalı f(t_f)
7	2	14	25
6	3	18	23
5	5	25	20
4	7	28	15
3	4	12	8
2	3	6	4
1	1	1	1
	n=25	**Σ fx =104**	

$$\text{Ortanca} = L + \left(\frac{n/2 - t_{fa}}{f_b}\right) a = 3.5 + \left(\frac{(12.5) - 8}{7}\right) 1 = 4,14$$ olarak hesaplanır.

Bu bulguya göre, grubun %50'sinin verilen metni 4,14 dakikadan daha az sürede, diğer yarısının ise daha uzun sürede okudukları söylenebilir.

ÖRNEK 2

n/2=100/2=50
50. gözlemin denk geldiği aralık için:
f_b : Frekans sayısı=13
t_{fa} : Öncesinin toplamalı frekansı= 41

Tablo 2.5: Başarı Testi Puanlarına Ait Gruplandırılmış Frekans Dağılımı

Puan Aralığı	f	rel.f	Orta Nokta	Gerçek sınırlar	Toplamalı f	Toplamalı rel.f
90-98	2	.02	94	89.5-98.5	100	1.00
81-89	6	.06	85	80.5-89.5	98	.98
72-80	9	.09	76	71.5-80.5	92	.92
63-71	12	.12	67	62.5-71.5	83	.83
54-62	17	.17	58	53.5-62.5	71	.71
45-53	13	.13	49	44.5-53.5	54	.54
36-44	15	.15	40	35.5-44.5	41	.41
27-35	14	.14	31	26.5-35.5	26	.26
18-26	8	.08	22	17.5-26.5	12	.12
09-17	4	.04	13	08.5-17.5	4	.04

$$\text{Ortanca} = L + \left(\frac{n/2 - t_{fa}}{f_b}\right)a = 44.5 + \left(\frac{50-41}{13}\right)9$$

Ortanca= 44.5 + 6.23 = 50.73’dir.

Ortalama, Ortanca ve Mod'un karşılaştırılması

<u>SAĞA ÇARPIK (POZİTİF ÇARPIK)</u>

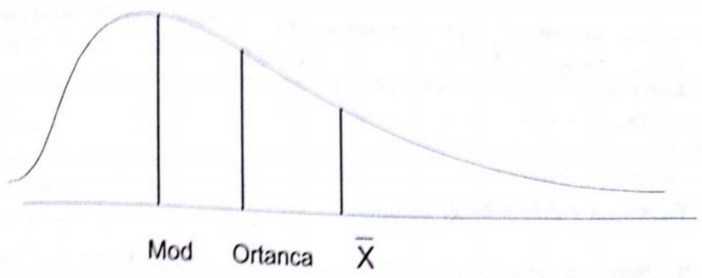
- $\bar{X}$ >Ortanca>Mod
- Ortalama, aşırı puanların bulunduğu sağ tarafa doğru çekilmekte ve böylece ortanca ortalamanın solunda kalmaktadır.
- Test puanları olarak düşünecek olursak, test öğrencilere _____ gelmiştir.
- Değerlerin çoğu ortalamanın altında

<u>SİMETRİK DAĞILIM</u>

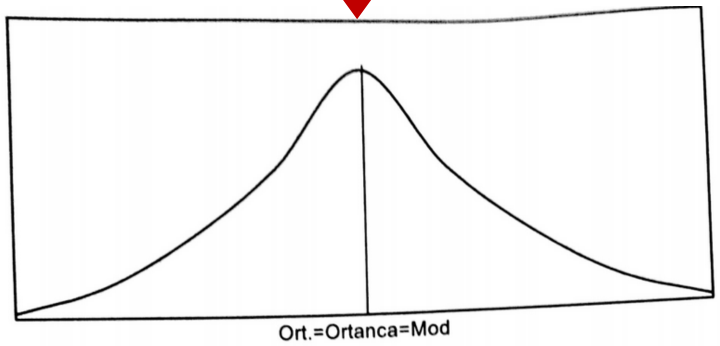
- Ortalama, mod ve medyan birbirine eşittir.
- Dağılım mükemmel simetriktir.
- Bu dağılımda en tutarlı ve kararlı merkezi eğilim ölçüsü, ortalamadır çünkü ortalama tüm puanlar dikkate alınarak hesaplanmaktadır ve ileri istatistikler için en uygun olandır.

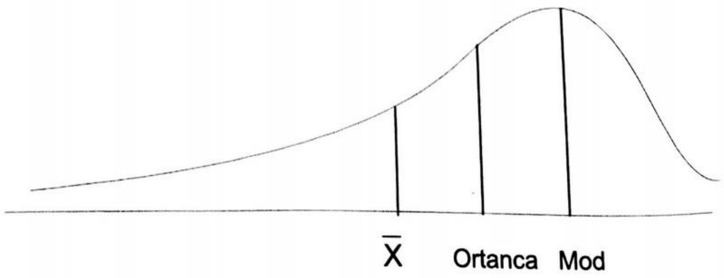
<u>SOLA ÇARPIK (NEGATİF ÇARPIK)</u>

- $\bar{X}$ <Ortanca < Mod
- Ortalama, aşırı puanların bulunduğu sol tarafa doğru çekilmekte ve böylece ortanca ortalamanın sağında kalmaktadır.
- Test puanları olarak düşünecek olursak, test öğrencilere _____ gelmiştir.
- Değerlerin çoğu ortalamanın üzerinde

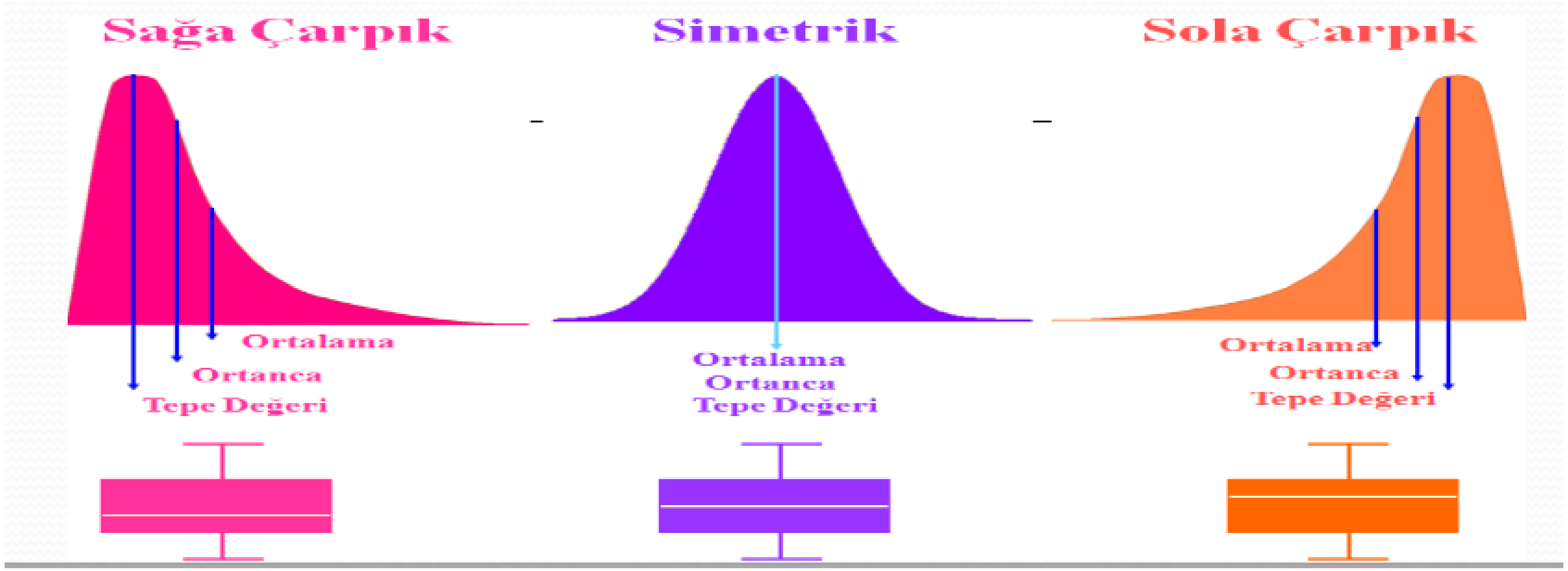


Şekil 3.3: Sağa Çarpık Dağılım





Şekil 3.2: Sola Çarpık Dağılım


Sağa Çarpık
Simetrik
Sola Çarpık
Ortalama
Ortanca
Tepe Değeri
Ortalama
Ortanca
Tepe Değeri
Ortalama
Ortanca
Tepe Değeri

Ortalama, Ortanca ve Mod’un Karşılaştırılması

- Aynı puan dağılımı için hesaplanan üç merkezî eğilim ölçüsü genellikle birbirinden farklıdır. Bu üç puana arasındaki büyük farklar, dağılımın simetrik olmadığını ya da bir yöne yatmış olduğunu gösterir.
- İstatistiksel analizler için dağılımların mükemmel bir şekilde simetrik olması gerekmez; ancak dağılımın olabildiğince normale yakın olması beklenmektedir (Howitt & Cramer, 1997).
- Ancak dağılım uç değerlerde belli bir birikimin sonucu olarak sağa ya da sola çarpık bir biçimde oluşuyorsa, uygun olan merkezi eğilim ölçüsü ortancadır çünkü ortalama puanların daha az gözlendiği çarpık tarafa doğru eğilim gösterirken ortanca bu çarpıklıktan etkilenmeyecektir(Büyüköztürk ve diğerleri, 2018).

…Ortalama, Ortanca ve Mod'un Karşılaştırılması…

a) Ortalama, ortanca ve moda oranla daha çok bilgiye dayanır. Çünkü, ortalama hesaplarken gözlenen ölçümlerin tümü işleme katılır. Halbuki, ortanca hesaplarken ölçümlerin yalnız yarısı, mod hesaplarken de biri ya da birkaçı işleme katılır. Ölçümlerin tümünün işleme katılması ortalamayı öbür vasat ölçülerinden daha istikrarlı ve güvenilir yapar. Başka bir deyişle, aynı evrenden seçilecek çeşitli örneklem grupları için hesaplanacak vasat ölçüleri arasında örneklemden örnekleme en az değişme gösteren ortalamadır.

...Ortalama, Ortanca ve Mod'un Karşılaştırılması...

b) Ortalama, vasat ölçüsü olmanın dışında başka matematiksel işlemler için çok daha uygundur; matematiksel işlemlere elverişlidir. Örneğin, gözlenen X lerin her birinin $\bar{X}$ dan olan farklarını (x) bulmak, bu farkların karelerini almak ve sonucu ölçümlerin birbirlerinden farklarının ölçüsü olarak kullanmak mümkündür. Daha sonraki bölümlerde göreceğimiz gibi, ortalamayı başka aritmetik işlemler için de kullanabiliriz. Öte yandan, ortanca ve mod, vasat ölçüsü olmanın dışında amaçlar için pek kullanılmaz.

…Ortalama, Ortanca ve Mod'un Karşılaştırılması

c) Ortalamanın değeri, ölçümlerin sayısal değerlerine bağlıdır. Bu nedenle, aşırı ölçümler ortalamanın değerini çok etkiler; ortalama aşırı ölçümlerin olduğu uca doğru çekilir. Dizideki ölçümlerden herhangi biri değişirse, ortalamanın değeri de değişir. Öte yandan, ortanca ve mod aşırı ölçümler tarafından çok az etkilenir. Dizinin tam ortasında bulunan ölçüm değişmeden ortancanın değeri, ölçümlerin dağılış biçimi değişmeden de modun değeri değişmez. Örneğin, yedi kişinin ağırlıkları 58, 58, 59, 60, 63, 64 ve 65 *kg*. olarak saptanmışsa, bu grup için $\bar{X} = 61$, *Ortn.* $= 60$ ve *Mod* $= 58$ olur. Söz konusu dizide ölçümlerin en büyüğü 65 değil de 79 olsaydı, $\bar{X} = 63$ olur, ortanca ve mod aynı kalırdı. Böylece, ortalamanın aşırı ölçümlere öbür vasat ölçülerinden daha duyarlı olduğu görülüyor. Bu uçtaki aşırı bir ölçüm ters uçta başka bir ölçüm tarafından dengelenmezse, ortalama aşırı ölçümün olduğu uca doğru çekilir. Bu durum özellikle gözlenen ölçüm sayısının az olduğu ve aşırı ölçümün diğerlerinden çok farklı olduğu durumlarda görülür.

Yüzdelik Hesaplama

- Yüzdelik ölçümlerin istenen bir yüzdesinin kendisinden aşağıda kaldığı değeri gösterir.
- Yüzdelik ölçek üzerinde, altında ve üstünde belirli oranları bulundurması istenilen noktanın değerine eşittir.
- Yüzdelik hesaplamasında aşağıdaki formül kullanılır;

$$Yy = L + \left(\frac{\frac{yn}{100} - t_{fa}}{f_b} \right) a$$

ÖRNEK

Tablo 2.5: Başarı Testi Puanlarına Ait Gruplandırılmış Frekans Dağılımı

Puan Aralığı	f	rel.f	Orta Nokta	Gerçek sınırlar	Toplamalı f	Toplamalı rel.f
90-98	2	.02	94	89.5-98.5	100	1.00
81-89	6	.06	85	80.5-89.5	98	.98
72-80	9	.09	76	71.5-80.5	92	.92
63-71	12	.12	67	62.5-71.5	83	.83
54-62	17	.17	58	53.5-62.5	71	.71
45-53	13	.13	49	44.5-53.5	54	.54
36-44	15	.15	40	35.5-44.5	41	.41
27-35	14	.14	31	26.5-35.5	26	.26
18-26	8	.08	22	17.5-26.5	12	.12
09-17	4	.04	13	08.5-17.5	4	.04

Örnek 3.6 Tablo 3.2'deki veriler için 45. yüzdeliği (Y_{45}) bulalım.

Çözüm. Çözüm için ilk aşama yn/100 değerini bulmaktır.

$\frac{yn}{100} = \frac{(45)(100)}{100}$ =45'dir. Bu değerin kapsandığı ilk toplamalı frekans 54 ve buna karşılık gelen puan aralığı 45-53'dür. Bu puan aralığı için bulunan L=44.5, t_{fa}=41, f_b=13 ve a=9 değerleri Formül 3.6'ya yerleştirilirse,

$$Y_{45} = L + \left(\frac{\frac{yn}{100} - t_{fa}}{f_b} \right) a = 44.5 + \left(\frac{45-41}{13}\right)9$$

Y_{45} = 44.5 + 2.77 = 47.27 bulunur.

Bu bulgu, grubun %45'nin 47.27 puanından düşük, kalan %55'nin ise yüksek puan aldığı şeklinde yorumlanır.

DEĞERLENDİRME

- Uç değerlerin bulunduğu bir veri setinde merkezî eğilim ölçüsü olarak hangisi tercih edilmelidir? Neden?
- Öğrencilerin sosyoekonomik düzeylere dağılımlarının alt-orta-üst olarak belirtildiği bir çalışmada hangi merkezî eğilim ölçüsü tercih edilmelidir?

KAYNAKLAR

Arıcı, H. (1998). *İstatistik: Yöntemler ve uygulama*. Kendi Yayını.

Baykul, Y. (1999). *İstatistik: Metodlar ve uygulamalar*. Ankara: Anı Yayıncılık.

Büyüköztürk, Ş., Çokluk, Ö. ve Köklü N. (2018). *Sosyal bilimler İçin istatistik*. Ankara: Pegem Akademi.

Howitt, D. & Cramer, D. (1997). *An introduction to statistics in psychology: A complete guide for students*. London: Prentice Hill.

Tan, Ş. (2016). *SPSS ve Excel Uygulamalı Temel İstatistik I.* Ankara: Pegem Akademi.