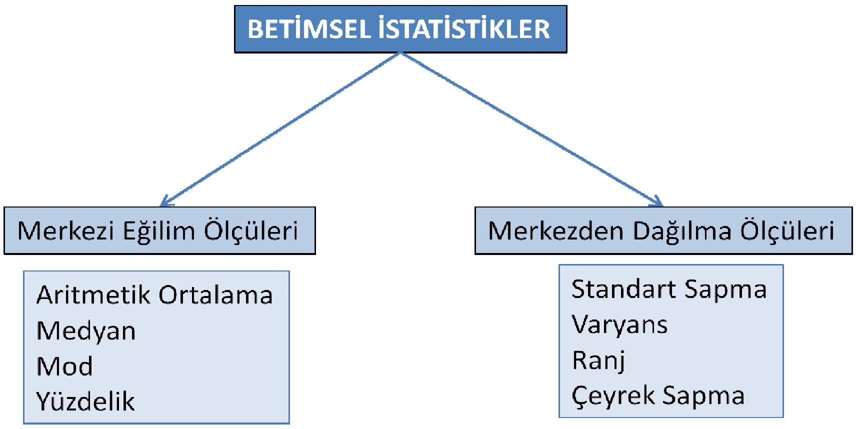
BÖLÜM 5

MERKEZİ EĞİLİM ÖLÇÜLERİ

Gözlenen belli bir özelliği, bu özelliğe ilişkin ölçme sonuçlarını yani verileri kullanarak betimleme, istatistiksel işlemlerin bir boyutunu oluşturmaktadır. Temel sayma ve sınıflama işlemleri ile elde edilen frekans tabloları ve grafik gösterimleri, betimlemenin yol ve yöntemlerinden biridir. Bunun yanı sıra 'betimsel istatistikler (descriptive statistics)' olarak adlandırılan bazı sayısal değerler kullanılarak da betimleme yapılabilmektedir.

Betimsel istatistikler, (1) merkezi eğilim ölçüleri ve (2) merkezden dağılma ölçüleri olmak üzere iki grupta sınıflandırılmaktadır:



Şekil 1. Betimsel İstatistiklerin Sınıflandırılması

Merkezi eğilim ölçüleri, 'merkeze yığılma ölçüleri' olarak da ifade edilebilmektedir. Aritmetik ortalama, mod, medyan, yüzdelik gibi nokta değerler, merkezi eğilim ölçüleridir. Diğer bir deyişle bu ölçüler, tek bir nokta belirtir. Bu nokta değerler verilerin yığılma noktaları olarak ilgilenilen özelliğe dönük betimlemelerin yapılmasında dikkate alınabilir.

Merkezi dağılım ölçüleri, 'merkezden yayılma ölçüleri' olarak da ifade edilebilmektedir. Standart sapma, varyans, ranj, çeyrek sapma gibi değerler, merkezden dağılma ölçüleridir. Bu ölçüler, merkez ya da ölçüt olarak belirlenen noktalara göre verilerin yayılması, çeşitlenmesi ya da farklılaşması hakkında bilgi verir.

Veri setinin karakteristiğine ve verilerin dağılımına göre uygun betimsel istatistiklerin kestirilmesi gerekir. Her betimsel istatistik her veri setinde anlamlı olmayabilir. Bu nedenle her bir betimsel istatistiğin hesaplanmasının yanı sıra hangi durumlarda kullanılabilir olduğunun da bilinmesi önemlidir. Aksi durumda elde edilen sayılar, yanıltıcı olabilir, yanlış ya da eksik yorumların yapılmasına yol açabilir.

Betimsel istatistikler tek başına, ilgilenilen özellik hakkında fazlaca bilgi sağlamaz. Birden fazla betimsel istatistik bir arada değerlendirilerek ya da birden fazla gruba/örnekleme yönelik betimsel istatistikler bir arada değerlendirilerek anlamlı betimsel yorumlar yapmak mümkündür.

Bu bölümde betimsel istatistiklerden merkezi eğilim ölçüleri grubunda yer alan 4 istatistik hakkında bilgi verilmektedir.

5.1. ARİTMETİK ORTALAMA[1]

Aritmetik ortalama (mean); her bir gözleme yönelik ölçme sonuçlarının toplamının gözlem sayısına bölünmesi ile elde edilen bir nokta değerdir. Aynı şekilde hesaplanmakla birlikte evren ortalaması ve örneklem ortalaması farklı sembollerle formülleştirilmektedir.

Evrenden elde edilen veriler üzerinde hesaplama yapılıyorsa, ortalama 'evren ortalaması (population mean)' olarak isimlendirilir. Evren ortalamasının formülü aşağıda verilmiştir:

$$\mu = \frac{x_1 + x_2 + \ldots + x_n}{N} = \frac{\sum x_i}{N}$$

Örneklemden elde edilen veriler üzerinde hesaplama yapılıyorsa ortalama, 'örneklem ortalaması (sample mean)' olarak isimlendirilir. Örneklem ortalamasının formülü aşağıda verilmiştir:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n}$$

[1] Aritmetik ortalama dışında, geometrik ortalama, harmonik ortalama gibi başkaca ortalama değerler de bulunmaktadır. Fakat bunlar içerisinde en bilindik olan ve en sık kullanılan aritmetik ortalamadır. Bu nedenle bundan sonraki bölümlerde aritmetik ortalama yerine 'ortalama' kullanımı tercih edilecektir. 'Ortalama' kullanımı, bundan sonraki bölümlerde, aritmetik ortalamayı ifade etmektedir.

Ortalama, normal dağılım gösteren ve en az eşit aralıklı ölçek düzeyinde olan sürekli verilerde, en 'sağlam' merkezi eğilim ölçüsü olarak bilinir. Bu varsayımları sağlayan verilerde ortalama, verilerin tamamını temsil eden bir sayısal değer olarak dikkate alınabilir.

Dikkat edilmesi gereken noktalardan biri, ortalama kestiriminin sürekli yani en az eşit aralıklı ölçek düzeyindeki değişkenler için anlamlı olduğudur. Cinsiyet, eğitim düzeyi, sınıf, şube, doğum yeri gibi kesikli verilerde ortalama, anlamlı bir istatistik değildir. Örneğin cinsiyet değişkeni için kız 1, erkek 2 olarak kodlanıp ortalama 1,63 olarak hesaplansa, bu sayıya karşılık gelen bir cinsiyet kategorisi bulunmadığı için bu sayı da anlamlı olmayacaktır.

Diğer bir önemli nokta ortalamanın, normal dağılım gösteren verilerde daha 'sağlam' bilgi vermesidir. Düşük ya da yüksek değerlere doğru yığılma gösteren yani çarpık dağılımlarda ortalama, yanıltıcı olabilmekte,yanlış ya da eksik yorumların yapılmasına yol açabilmektedir.

Ortalama, uç noktalardan aşırı etkilenen bir istatistiktir. *Uç nokta, grubun ya da örneklemin genelinden manidar düzeyde ayrışan ölçme sonuçlarını ifade etmektedir.* Örneğin bir yazılı yoklamada 2 öğrenci 40'ın altında diğer öğrenciler ise 50'nin üzerinde notlar almışsa, 40'ın altındaki bu notlar uç nokta oluşturur ve sınıf ortalamasını aşağıya çeker. Normal dağılım gösteren veriler, uç noktalardan, manidar düzeyde etkilenmez. Bu nedenle ortalamanın kullanılmasında verilerin dağılımına dikkat edilmesi gerekir.

ÖRNEK 1.

12 öğrencinin günlük bilgisayar kullanım süreleri şu şekildedir:

2sa; 3sa; 1sa; 1,5sa; 1sa; 4sa; 2,5sa; 2sa; 3sa; 2sa; 1,5sa; 2sa

Buna göre bu öğrencilerin günlük ortalama bilgisayar kullanma süresini hesaplayalım. Bunun için ölçme sonucu olarak verilen 12 sayıyı toplayıp 12'ye bölmemiz gerekir.

$$\bar{X} = \frac{2 + 3 + 1 + 1{,}5 + 1 + 4 + 2{,}5 + 2 + 3 + 2 + 1{,}5 + 2}{12}$$

$$= \frac{25{,}5}{12}$$

$$= 2{,}125$$

Bu 12 öğrencinin günlük bilgisayar kullanma sürelerinin ortalaması 2,125 saattir. Yukarıda açıklanan normal dağılım gösterme koşulunun sağlanabilmesi için daha fazla sayıda veriye ihtiyaç vardır. Normal dağılımın sağlandığı bir örnekte bu istatistik "öğrenciler, günlük ortalama 2,125sa bilgisayar kullanmaktadır" şeklinde bir yorumun yapılmasını mümkün kılar.

5.2. MEDYAN (ORTANCA)

Medyan (median); küçükten büyüğe doğru sıralanmış verilerin tam ortasında kalan değerdir. Medyan, sıralanmış verileri %50 %50 olarak ikiye bölen noktadır ve grubun yarısı hakkında bilgi verir.

Medyan, sıralama işlemine dayalı olduğu için en az sıralama ölçeği düzeyindeki değişkenlerde anlamlı bir betimsel istatistiktir. Cinsiyet, şube, doğum yeri, okul türü gibi sınıflama ölçeği düzeyindeki kategorik değişkenlerde medyan anlamlı değildir. Bu tür değişkenlerde kategoriler arasında bir sıra ilişkisi bulunmamaktadır.

Sürekli verilerde ortalama, medyana göre daha 'sağlam' bir istatistiktir. Bununla birlikte değişkene yönelik verilerin dağılımı normal dağılımdan sapma gösterdiği ya da uç noktaların etkisinin olduğu durumlarda ortalama yanıltıcı olur. Bu durumda ortalama yerine medyanın dikkate alınması önerilir.

ÖRNEK 2.

a) Veri Sayısı Tek Olduğunda

Bir sınıftaki 25 öğrencinin bireysel kitaplığında bulunan kitap sayıları şu şekildedir:

12; 15; 10; 8; 12; 16; 20; 47; 22; 16; 18; 15; 16; 19; 23; 45; 20; 16; 10; 15; 7; 20; 16; 10; 11

Önce, daha kolay okuyabilmek için verileri sıralayalım. Sıralı veriler aşağıda verilmiştir:

7; 8; 10; 10; 10; 11; 12; 12; 15; 15; 15; 16; 16; 16; 16; 16; 18; 19; 20; 20; 20; 22; 23; 45; 47

Verilerde uç nokta olarak 45 ve 47 değerleri dikkat çekmektedir. Bunların altında yer alan en yüksek değer 23'tür. Arada 22 puanlık bir fark var. Yukarıda açıklandığı gibi bu uç noktalar ortalamayı yukarı çekecek ve yanıltıcı olacaktır. Bunu görmek için hem 25 verinin ortalamasını hem bu iki veri ihmal edildiğinde kalan 23 verinin ortalamasını hesaplayalım:

$$\overline{X}_1 = \frac{7 + 8 + 10 + 10 + \ldots + 23 + 45 + 47}{25} = \frac{439}{25} = 17,56$$

$$\overline{X}_2 = \frac{7 + 8 + 10 + 10 + \ldots + 22 + 23}{23} = \frac{347}{23} = 15,09$$

Görüldüğü gibi uç noktalar ortalamanın manidar düzeyde yükselmesine neden olabilmektedir. Bu durumda ortalama yerine medyan kullanılması daha doğru olacaktır.

Veri setinde 25 veri olduğuna göre sıralama yapıldıktan sonra 13. sıradaki değer medyanı verecektir.

7; 8; 10; 10; 10; 11; 12; 12; 15; 15; 15; 16; **16**; 16; 16; 16; 18; 19; 20; 20; 20; 22; 23; 45; 47

13. sıra

Kitap sayısını gösteren ölçme sonuçlarının medyanı 16'dır. Bu değer grubun yarısı hakkında bilgi verir. Yani "öğrencilerin yarısının bireysel kitaplığında 16 ve üzerinde kitap bulunmaktadır" ya da "öğrencilerin yarısının kitaplığında 16 ve altında kitap bulunmaktadır" şeklinde betimsel yorumlar yapılabilir.

b) Veri Sayısı Çift Olduğunda

Yukarıdaki örnekte 25 öğrenci bulunmaktaydı. Bu durumda sıralanmış verilerin tam ortasında 13. sıradaki veri kalmaktadır. Peki veri sayısı çift sayı olsaydı! Örneğin kitap sayıları aşağıdaki gibi 24 öğrenciye yönelik olsaydı:

15,5

Kitap Sayısı: 0; 2; 10; 10; 10; 11; 12; 12; 13; 14; 15; **15**; **16**; 16; 16; 16; 18; 19; 20; 20; 20; 22; 23; 25

12. sıra

13. sıra

Bu durumda ortada iki sayı değeri kalmaktadır: 12 ve 13. sıradaki sayılar. 12. sıradaki değer 15, 13. sıradaki değer 16'dır. Medyan bu iki değerin ortalamasıdır. O halde medyan 15,5 olarak elde edilir:

$$\text{Medyan} = \frac{15 + 16}{2}$$

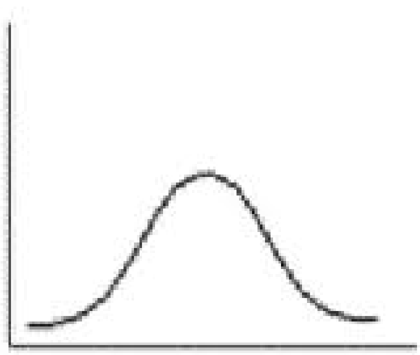
$$= 15,5$$

5.3. MOD (TEPE NOKTASI)

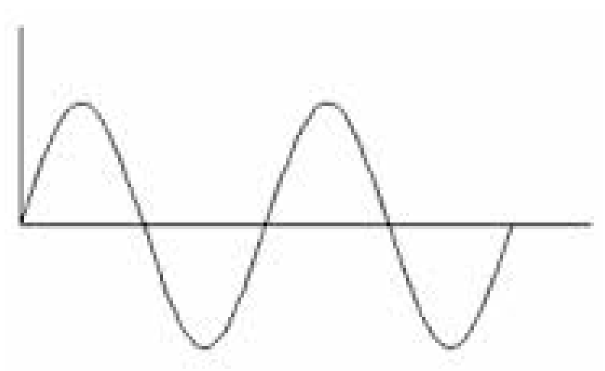
Mod, frekansı yani sıklığı en fazla olan değerdir. Sütun grafiği, histogram ya da çizgi grafiği gibi grafiklerde, dağılımın maksimum noktasını ifade eden frekans değeri, mod olarak belirlenir.

Mod, sınıflama ölçeği düzeyindeki değişkenlerde dahil olmak üzere her tür değişkende anlamlı bir betimsel istatistiktir.

Bir veri setinde görülme sıklığı en yüksek olan bir değer bulunabileceği gibi birden fazla değer için de görülme sıklığı eşit ve en yüksek olabilir. Mod olarak tek bir değerin belirlenebildiği değişkenlere 'tek modlu değişken' denir ve bu değişkenin değerlerinin dağılımı tek tepeli bir dağılım gösterir. Birden fazla mod değeri olan değişkenlere ise 'çok modlu değişken' denir ve bu değişkenlerin gözlenen değerlerinin dağılımı çok tepeli dağılım gösterir.



Grafik 1. Normal Dağılım Eğrisi



Grafik 2. Sinüs Eğrisi

ÖRNEK 3.

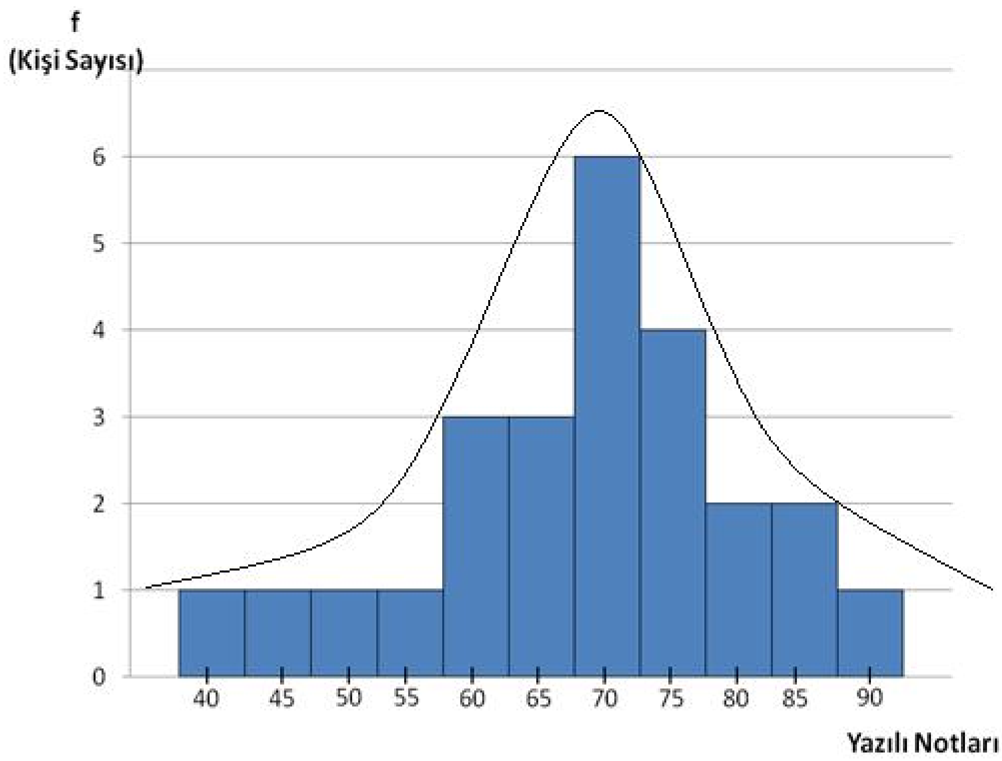
Bir sınıftaki 25 öğrencinin yazılı sınav notları aşağıda verilmektedir:

NOT: 55; 60; 65; 75; 90; 95; 90; 80; 75; 75; 70; 65; 60; 50; 45; 40; 70; 65; 70; 70; 60; 70; 80; 75; 70

SIRALI NOTLAR: 40; 45; 50; 55; 60; 60; 60; 65; 65; 65; 70; 70; 70; 70; 70; 70; 75; 75; 75; 75; 80; 80; 85; 85; 90

Görüldüğü gibi en sık tekrar eden ya da frekansı en yüksek olan not 70'dir. 70 değerinin frekansı 6'dır. Bu durumda yazılı notlarının modu 70 olur.

Yazılı notlarının histogramı çizildiğinde de mod yani tepe noktasının 70 olduğu açık bir şekilde görülmektedir. Söz konusu histogram aşağıda verilmektedir:



Grafik 3. Yazılı Notlarının Dağılımı

Eğer veri setinde birden fazla mod değeri olsaydı örneğin 70 notu 6 defa gözlenirken, 60 notu ve 85 notu da 6 defa gözlenmiş olsaydı, bu değişken çok modlu bir değişkendir denilirdi. Dağılımın çok tepeli olması, normal dağılımdan sapma olduğunun bir göstergesidir. Bu durumda normal dağılım göstermesi durumunda kullanılacak istatistikler ve istatistiksel yönetmelerin kullanılması yanıltıcı olabilir.

5.4. YÜZDELİK VE ÇEYREK SAPMA

Yüzdelik; küçükten büyüğe doğru sıralanmış verilerin belli bir yüzdesini altında bırakan noktadaki gözlenen değerdir. '$Y_{yüzde}$' sembolü ile gösterilir. Örneğin;

- Y_{20}; sıralanmış verilerin %20'sini aşağısında bırakan sıradaki gözlenen değeri,
- Y_{25}; sıralanmış verilerin %25'ini aşağısında bırakan sıradaki gözlenen değeri,
- Y_{50}; sıralanmış verilerin %50'sini aşağısında bırakan sıradaki gözlenen değeri,
- Y_{75}; sıralanmış verilerin %75'ini aşağısında bırakan sıradaki gözlenen değeri,

ifade eder.

Çeyreklik (quartile) ise sıralanmış verilerin çeyrek yani %25'lik dilimlerinin başlangıç ve bitiş noktalarında yer alan değerleri ifade eder. Dolayısıyla çeyreklik değerler aynı zamanda yüzdelik değerlerle ifade edilebilir. Çeyreklik, 'Q' sembolü ile ifade edilir.

Teorik olarak, sıralanmış bir veri setinde 4 çeyrek bölüm vardır. Dördüncü çeyreklik, verilerin %100'ünün yani tamamının bir üstündeki değeri ifade eder. Bu nedenle pratikte dördüncü çeyreklik kullanılmaz. Birinci çeyreklik Q_1, ikinci çeyreklik Q_2 üçüncü çeyreklik Q_3 sembolleri ile gösterilir.

Birinci çeyreklik 25. yüzdelik ile ikinci çeyreklik 50. yüzdelik ile üçüncü çeyreklik 75. yüzdelik ile örtüşür.

$Q_1=Y_{25}$

$Q_2=Y_{50}$

$Q_3=Y_{75}$

Çeyreklik değerlerin hesaplanmasında hangi sıradaki gözlenen değerin çeyreklik olduğunun belirlenmesinde aşağıdaki formüller kullanılabilir:

$$Q_1 = \frac{n+1}{4} \text{ sıradaki değer}$$

$$Q_2 = 2.\frac{n+1}{4} \text{ sıradaki değer}$$

$$Q_3 = 3.\frac{n+1}{4} \text{ sıradaki değer}$$

Yüzdelik ve çeyreklik, en az eşit aralıklı ölçek düzeyindeki değişkenlerde anlamlı bir istatistiktir. Cinsiyet, medeni durum, şube gibi sınıflama ölçeği düzeyindeki değişkenlerde bu istatistik anlamlı değildir. Eğitim düzeyi, sınıf gibi sıralama ölçeği düzeyindeki değişkenlerde ise yüzdelik ve çeyreklik yerine 'sıra farkları (rank)' kullanılmaktadır.

ÖRNEK 4.

Örnek 2'de verilen öğrencilerin bireysel kitaplıklarında bulunan kitap sayısı gözlem değerleri için yüzdelik ve çeyreklik değerler hesaplaması yapalım. Bunun için söz konusu 25 gözlem değeri aşağıda küçükten büyüğe doğru sıralanmış olarak verilmiştir.

7; 8; 10; 10; 10; 11; 12; 12; 15; 15; 15; 16; 16; 16; 16; 16; 18; 19; 20; 20; 20; 22; 23; 45; 47

a) 10. yüzdelik değerini hesaplayalım.

10. yüzdelik, sıralanmış verilerin %10'unu aşağısında bırakan sıradaki gözlenen değer olacaktır. 25 veri olduğuna göre 25'in %10'u 2,5 olarak hesaplanır. O halde 10. yüzdelik 2,5 veriyi aşağıda bırakan sıradaki yani üçüncü sıradaki değer olacaktır. O değer 10'dur.

7; 8; **10**; 10; 10; 11; 12; 12; 15; 15; 15; 16; 16; 16; 16; 16; 18; 19; 20; 20; 20; 22; 23; 45; 47

$\boxed{Y_{10}}$

b) 25. yüzdelik yani birinci çeyreklik değerini hesaplayalım.

25. yüzdelik, sıralanmış verilerin %25'ini aşağısında bırakan sıradaki gözlenen değer olacaktır. 25 veri olduğuna göre 25'in %25'i 6,25 olarak hesaplanır. O halde 25. yüzdelik 6,25 veriyi aşağıda bırakan sıradaki yani yedinci sıradaki değer olacaktır. O değer 12'dir.

7; 8; 10; 10; 10; 11; **12**; 12; 15; 15; 15; 16; 16; 16; 16; 16; 18; 19; 20; 20; 20; 22; 23; 45; 47

$\boxed{Y_{25}}$

c) 40. yüzdelik değerini hesaplayalım.

40. yüzdelik, sıralanmış verilerin %40'ını aşağısında bırakan sıradaki gözlenen değer olacaktır. 25 veri olduğuna göre 25'in %40'ı 10 olarak hesaplanır. O halde 40. yüzdelik 10 veriyi aşağıda bırakan sıradaki yani on birinci sıradaki değer olacaktır. O değer 20'dir.

7; 8; 10; 10; 10; 11; 12; 12; 15; 15; **15**; 16; 16; 16; 16; 16; 18; 19; 20; 20; 20; 22; 23; 45; 47

$\boxed{Y_{40}}$

d) 75. yüzdelik yani üçüncü çeyreklik değerini hesaplayalım.

75. yüzdelik, sıralanmış verilerin %75'ini aşağısında bırakan sıradaki gözlenen değer olacaktır. 25 veri olduğuna göre 25'in %75'i 18,75 olarak hesaplanır. O halde 75. yüzdelik 18,75 veriyi aşağıda bırakan sıradaki yani on dokuzuncu sıradaki değer olacaktır. O değer 20'dir.

7; 8; 10; 10; 10; 11; 12; 12; 15; 15; 15; 16; 16; 16; 16; 16; 18; 19; **20**; 20; 20; 22; 23; 45; 47

$\boxed{Y_{75}}$