

# VERİLERİN ÖZETLENMESİ

BİYOİSTATİSTİK ve TIP BİLİŞİMİ A.D.

# İSTATİSTİĞİN ÜÇ TEMEL GÖREVİ

İstatistik genelde üç temel görev üstlenir. Bunlar:

**1. Tanımlama (description)**

2. Analiz

3. Tahmin(Prediction)

❖ Tanımlayıcı (decriptive) istatistik olaylar veya eşyalar hakkında toplanmış kalabalık sayı yığınlarını özetleyip onları daha kolay anlaşılır hale getirir. Bunlar daha ziyade özet istatistikler ve grafikler yardımı ile yapılır.

# NİTEL (KALİTATİF) DEĞİŞKEN NEDİR?

❖ Adlandırma ölçeğinde değişken (cinsiyet, kan grubu, hastalık türü vb) değişkenler nitel değişkenlerdir.

❖ Örneğin:

❖ Kan bağışı yapan 25 kişinin kan grupları (**kalitatif** bir özellik) aşağıdaki şekilde belirleniyor. (Verinin **gruplandırılmamış** hali)

AB	B	A	O	B
O	B	O	A	O
B	O	B	B	B
A	O	AB	AB	O
A	B	AB	B	A

Bu tip bir değişkeni nasıl özetleyip istatistik yöntemlerle değerlendirme yapabiliriz?

Sonuçlar tekrar düzenlenerek her **kan grubunun** ortaya çıkış sıklığı elde edilirse, bu düzenlenmiş tabloya **frekans tablosu** adı verilir (verinin **gruplandırılmış** hali).

## GRUPLANDIRILMIŞ VERİ

**Tablo:- Örneğe ait frekans tablosu**

<b>Kan Grubu</b>	<b>Frekans (Sıklık)</b>
<b>A</b>	<b>5</b>
<b>B</b>	<b>9</b>
<b>O</b>	<b>7</b>
<b>AB</b>	<b>4</b>
<b>TOPLAM</b>	<b>25</b>

Bu tip frekans tablosundan yapılabilecek hesaplamalar nelerdir?

Kan Grubu	Frekans	Eklemeli frekans	Göreceli frekans	Eklemeli Göreceli Frekans
A	5	5	0,200	0,200
B	9	14	0,360	0,560
O	7	21	0,280	0,840
AB	4	25	0,160	1,000
<b>TOPLAM</b>	<b>25</b>		<b>1,000</b>	

$$\text{Göreceli Frekans}_{(i \text{ nci grup})} = \frac{f_i}{\sum_{i=1}^k f_i}$$

İstenirse göreceli frekanslar 100 ile çarpılarak % değerler elde edilir.

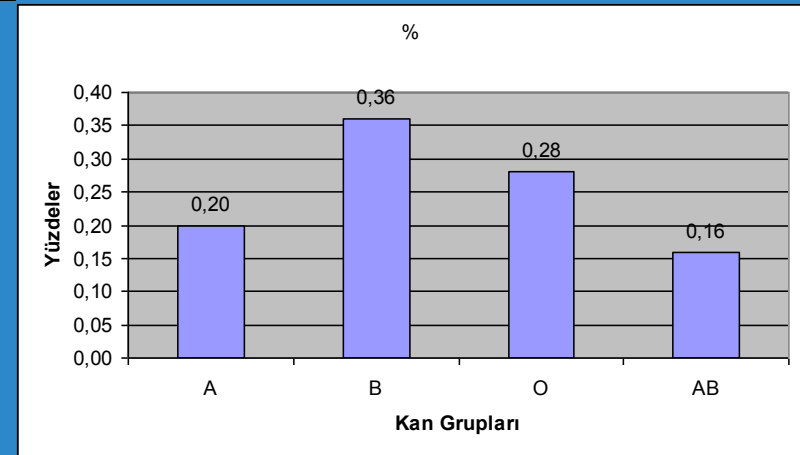
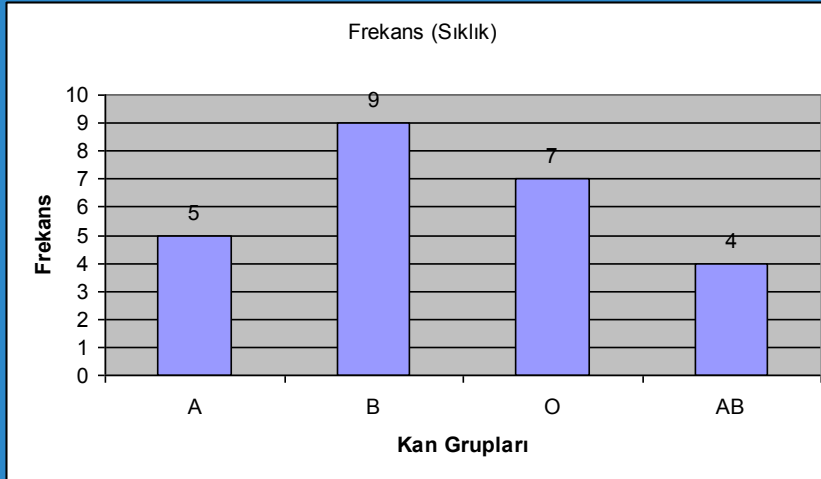
$$\text{Göreceli Frekans}_{(A)} = p_a = \frac{5}{25} = 0.200$$

$$\text{Göreceli Frekans}_{(B)} = p_b = \frac{9}{25} = 0.360$$

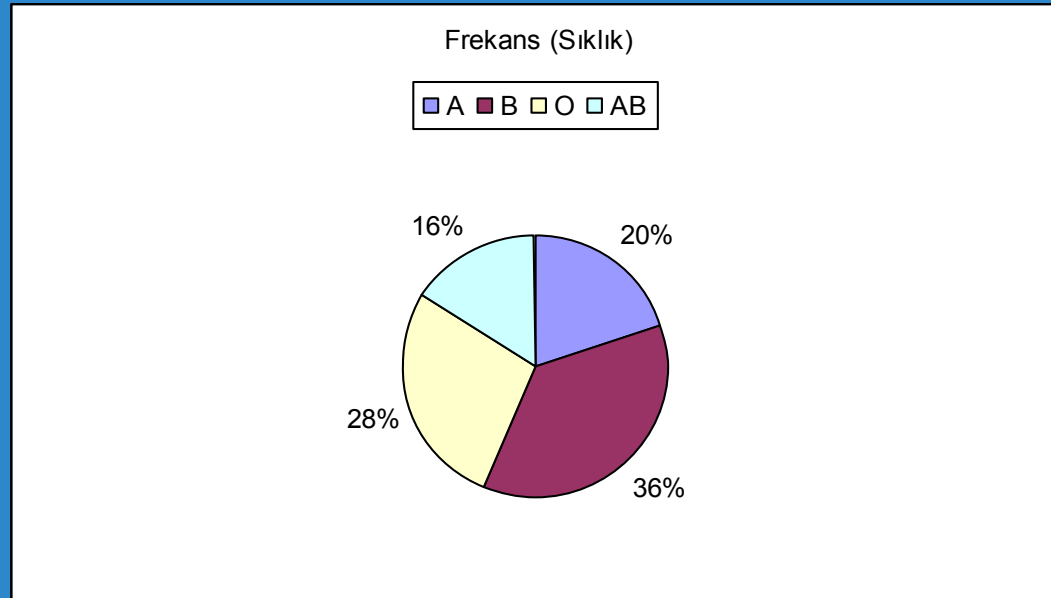
$$\text{Göreceli Frekans}_{(O)} = p_o = \frac{7}{25} = 0.280$$

$$\text{Göreceli Frekans}_{(AB)} = p_{ab} = \frac{4}{25} = 0.160$$

# SÜTUN GRAFİKLER (FREKANS),(GÖRECELİ FREKANS)



# PASTA GRAFİK



Pasta Grafik üzerinde frekanslar veya yüzde değerler verilebilir.



# NİCEL (KANTİTATİF) DEĞİŞKENLER

Yeni doğan çocukların ağırlıkları ile ilgilenen bir araştırmada çocuk doğum ağırlıkları oransal ölçekte bir nicel değişkendir.

Böyle bir değişkeni sütun grafikte göstermek için değerlerin sınıflandırılması veya gruplandırılması gerekir. Bunu yapmak için kaç gruba ayrılacağı belirlenmeli, sınıf aralıkları belirlenmeli, sonra da gruplar oluşturulmalıdır.

# FREKANS(SIKLIK) TABLOSU

Yüz yeni doğan çocuğun doğum ağırlıkları (kg) aşağıdaki gibi bulunmuştur.

2,6	2,6	3,0	2,7	3,8	2,7	3,1	2,7	4,8	4,9
3,6	4,0	4,8	2,7	3,9	2,4	2,0	3,4	2,0	3,9
3,6	4,7	2,3	2,4	3,7	2,7	1,7	2,9	2,4	2,1
5,1	3,2	3,1	4,1	3,0	4,3	3,8	3,3	2,4	3,2
1,7	5,0	4,4	2,6	3,0	3,6	4,7	2,6	2,8	2,7
3,1	2,0	2,4	3,5	2,2	3,9	2,9	3,6	2,6	4,0
2,4	3,3	4,1	2,4	2,2	3,1	4,2	3,9	2,4	3,3
3,8	3,5	5,2	2,5	3,4	3,2	3,8	4,1	2,7	3,7
3,3	4,3	3,4	2,1	2,9	3,8	2,2	3,2	2,8	2,9
2,7	5,0	3,1	2,9	2,7	3,5	2,0	2,5	4,7	2,5

❖ **Kaç sınıf** yapılacağını yaklaşık hesaplamak için **Sınıf Sayısı=1 + 3.322 · log<sub>10</sub>(N)** formülü kullanılabilir. Buradan da **sınıf genişliği** hesaplanabilir.

$$C = \frac{X_{\text{enbüyük}} - X_{\text{enküçük}}}{1 + 3.322 \cdot \log(N)}$$

$$\text{Sınıf Genişliği} = \frac{\text{Değişim Aralığı}}{\text{Sınıf Sayısı}}$$

# SINIF SAYISI VE SINIF GENİŞLİĞİNİ BELİRLEMEK

Örneğe bu formül uygulansaydı:

$$\text{Sınıf sayısı} = 1 + 3.322 \cdot (\log 100) = 1 + 6,644$$

$= 7.644 \cong 8$  sınıf yapılabilir.

Buna göre **sınıf genişliği** ne olacak?

$$X_{\text{enbüyük}} = 5.2, X_{\text{enküçük}} = 1.7$$

$$C = (5.2 - 1.7) / 8 = 0.4375 \cong 0.5 \text{ alınabilir.}$$

# KURALLAR

Sınıflar oluşturulurken belirli kurallara uymak gerekir.

❖ Çok fazla sayıda sınıf yapılırsa veri **iyi** **özetlenmemiş** olur.

❖ Az sınıf yapılırsa **bilgi kaybı** çok fazla olur.

Bu nedenle optimum sayıda sınıf oluşturmalıdır.

Yani:

# **TABLO OLUŞTURURKEN DİKKAT EDİLECEK HUSUSLAR**

1. Oluşturulan sınıflar **verilerin hepsini kapsamalıdır**. En **küçük gözlem ilk sınıfa**, en **büyük gözlem de en son sınıfa** düşmelidir.
2. **Sınıflar arası boşluklara** hiç bir gözlem düşmemelidir.
3. **Her gözlem sadece bir sınıfa** düşmelidir. Yani sınıfların sınırları birbiri ile **çakışmamalıdır**.
4. Mümkün olduğu **kadar sınıf genişlikleri tüm sınıflar için eşit** olmalıdır. Bu karşılaştırma yapılırken önem kazanır.

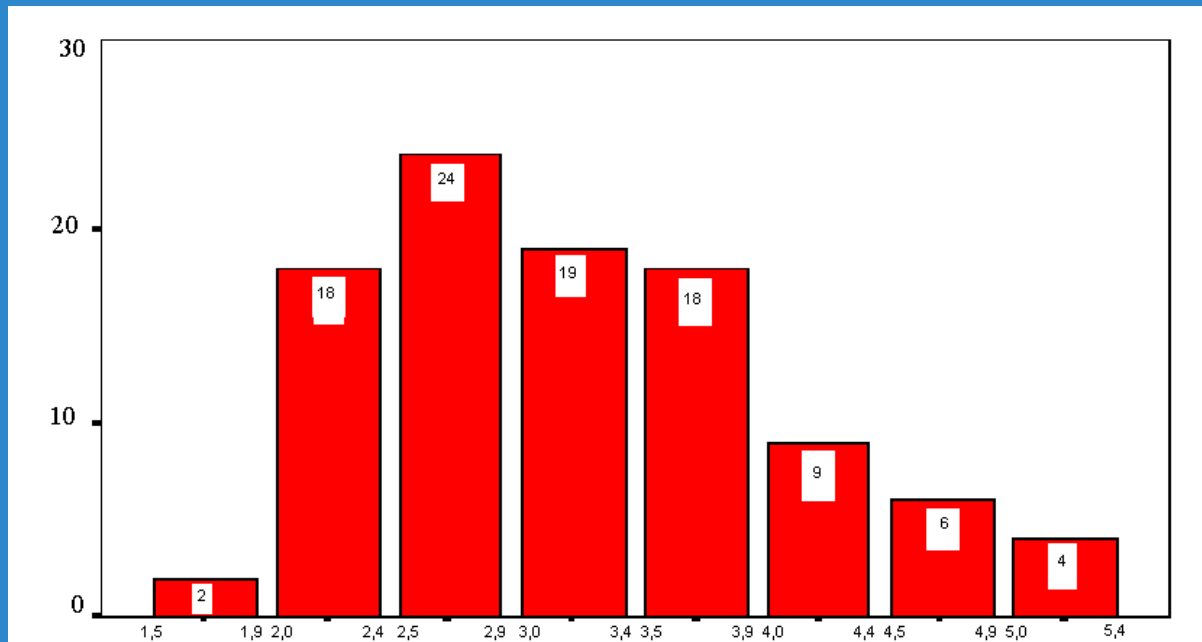
❖ Sınıf genişliği **0.5** kg olacak şekilde doğum ağırlıklarını sıklık(frekans) tablosu halinde özetlemek için **8** sınıf yapılabilir.

<b>Sınıf limitleri</b>	<b>Sınıf sınırları</b>	<b>Sıklık(frekans)</b>
<b>1.5 - 1.9</b>	<b>1.45 - 1.94</b>	<b>2</b>
<b>2.0 - 2.4</b>	<b>1.95 - 2,44</b>	<b>18</b>
<b>2.5 - 2.9</b>	<b>2,45 – 2,94</b>	<b>24</b>
<b>3.0 - 3.4</b>	<b>2,95 – 3,44</b>	<b>19</b>
<b>3.5 - 3.9</b>	<b>3,45 – 3,94</b>	<b>18</b>
<b>4.0 - 4.4</b>	<b>3,95 – 4,44</b>	<b>9</b>
<b>4.5 - 4.9</b>	<b>4,45 – 4,94</b>	<b>6</b>
<b>5.0 - 5.4</b>	<b>4,95 – 5,44</b>	<b>4</b>
<b>Toplam</b>		<b>100</b>

# Sınıfların orijinal kayıt sınırları (sınıf limitleri)

(İlk sınıf: 1.5-1.9..... Son sınıf: 5.0-5.4)

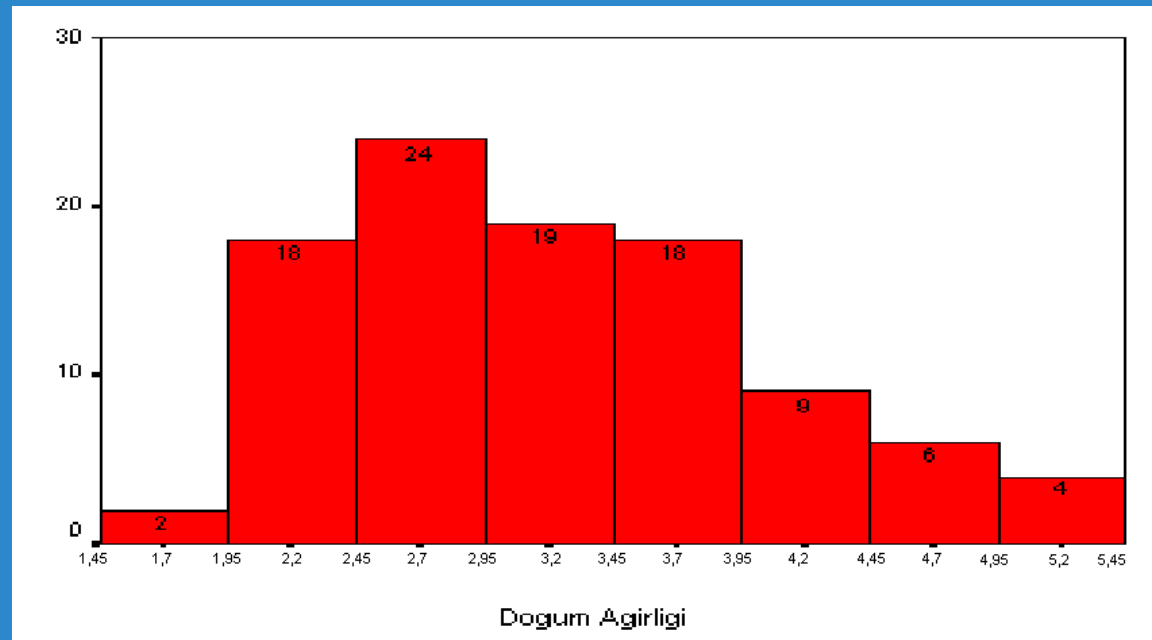
## Çubuk Grafik



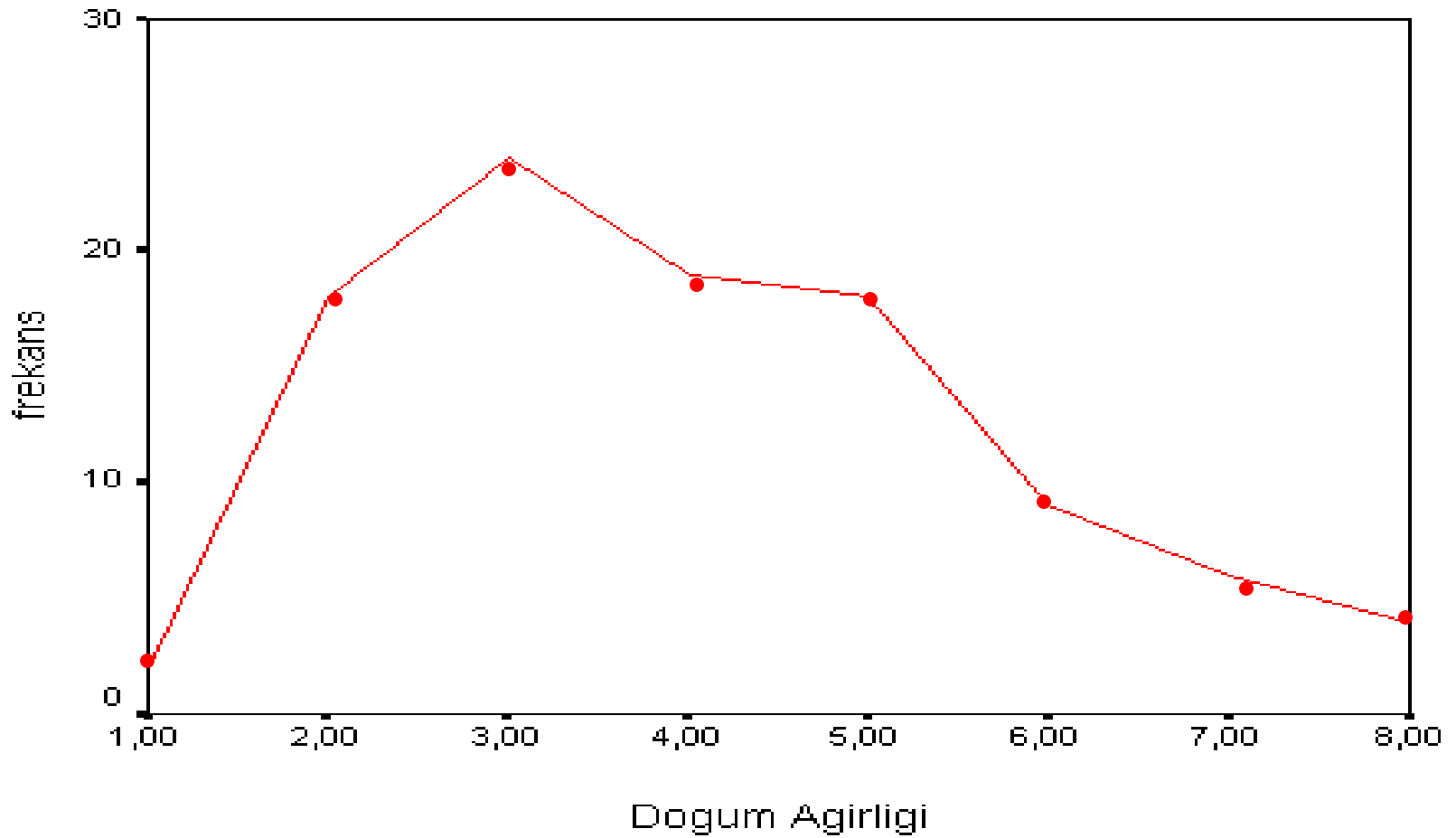


# Ayarlanmış sınırlar(sınıf sınırları) Histogram

(İlk sınıf: 1.45-1.94.... Son sınıf:4.95-5.44)



# Çocuk Doğum Ağırlığı Frekans Poligonu



Daire grafik çizmek için nisbi (göreceli) frekanslar bulunur ve bir dairenin alanı bu oranlarda parçalara bölünerek grafik çizilir.

Bunu yapmak için her parça için kaç derecelik açı ayrılacağı bulunur, bunun için

$a_i = 360 * f_i / N$  formülü kullanılarak her sınıf için ayrılacak açı değeri bulunur.

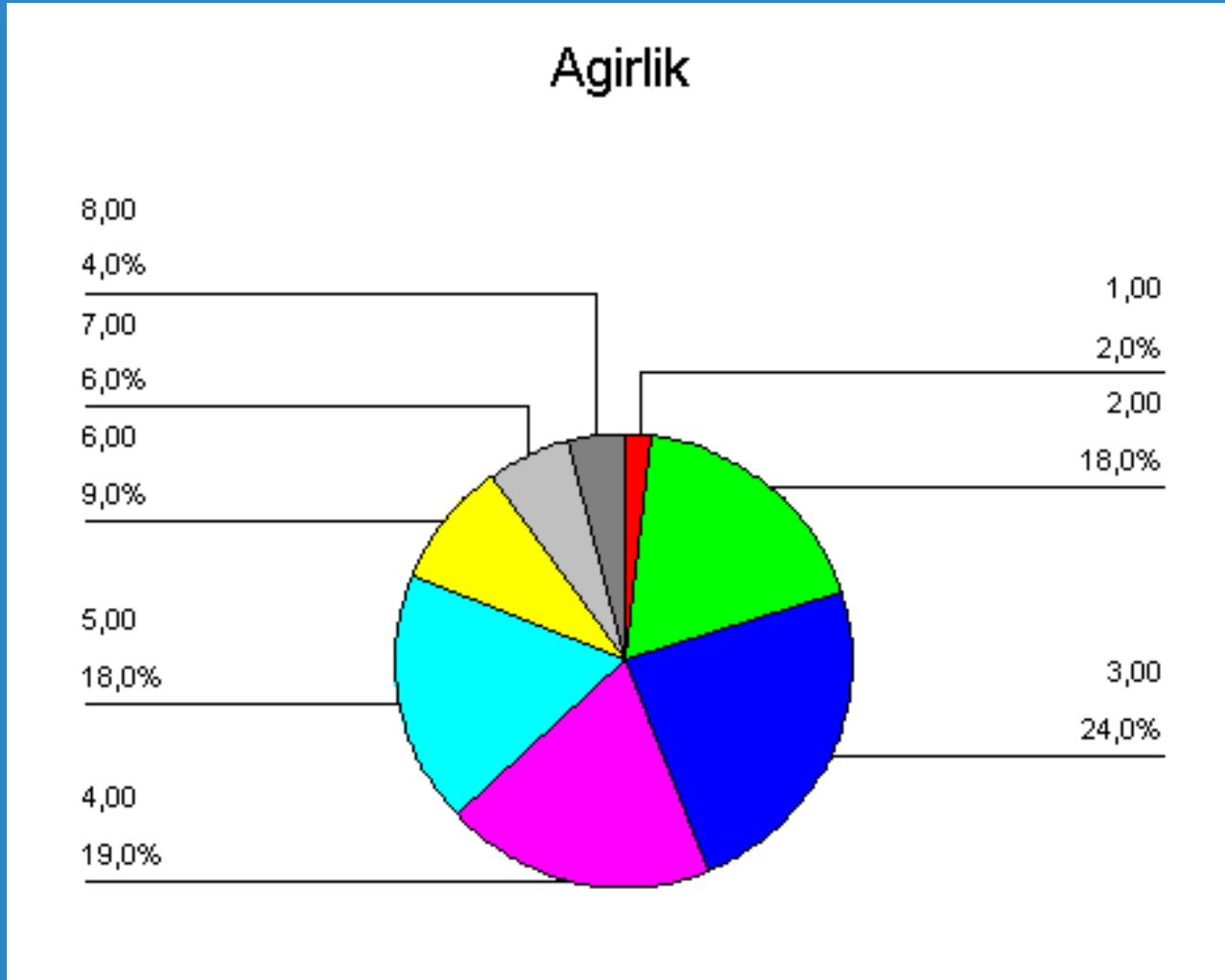
$$a_1 = 360 * 2 / 100 = 7.2 \text{ derece}$$

$$a_2 = 360 * 18 / 100 = 64.8 \text{ derece}$$

Aynı şekilde diğerleri de hesaplanır.

$$a_8 = 360 * 4 / 100 = 14.4 \text{ derece}$$

# ÇOCUK AĞIRLIĞI İÇİN DAİRE GRAFİK

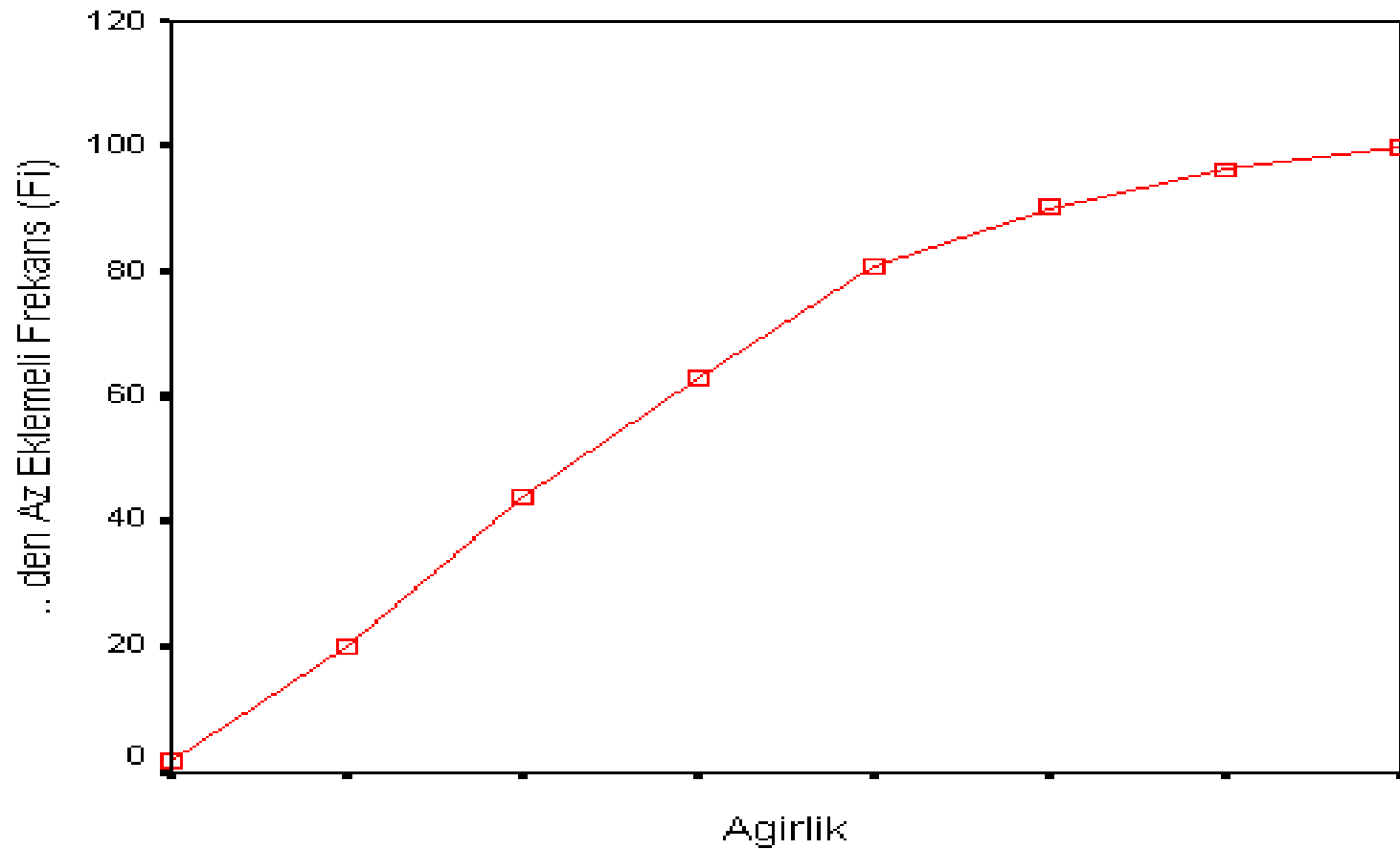


# EKLEMELİ SIKLIKLAR (..DEN AZ **Fi**)

(1.94 DEN AZ AĞIRLIĞI OLAN 2 ÇOCUK VAR)

Sınıf sınırları	Sıklık (frekans= $f_i$ )	....den Az ( $F_i$ )
1.45 - 1.94	2	2
1.95 - 2,44	18	20
2,45 – 2,94	24	44
2,95 – 3,44	19	63
3,45 – 3,94	18	81
3,95 – 4,44	9	90
4,45 – 4,94	6	96
4,95 – 5,44	4	100
	100	

# EKLEMELİ SIKLIKLAR (..DEN AZ $F_i$ ) GRAFİĞİ

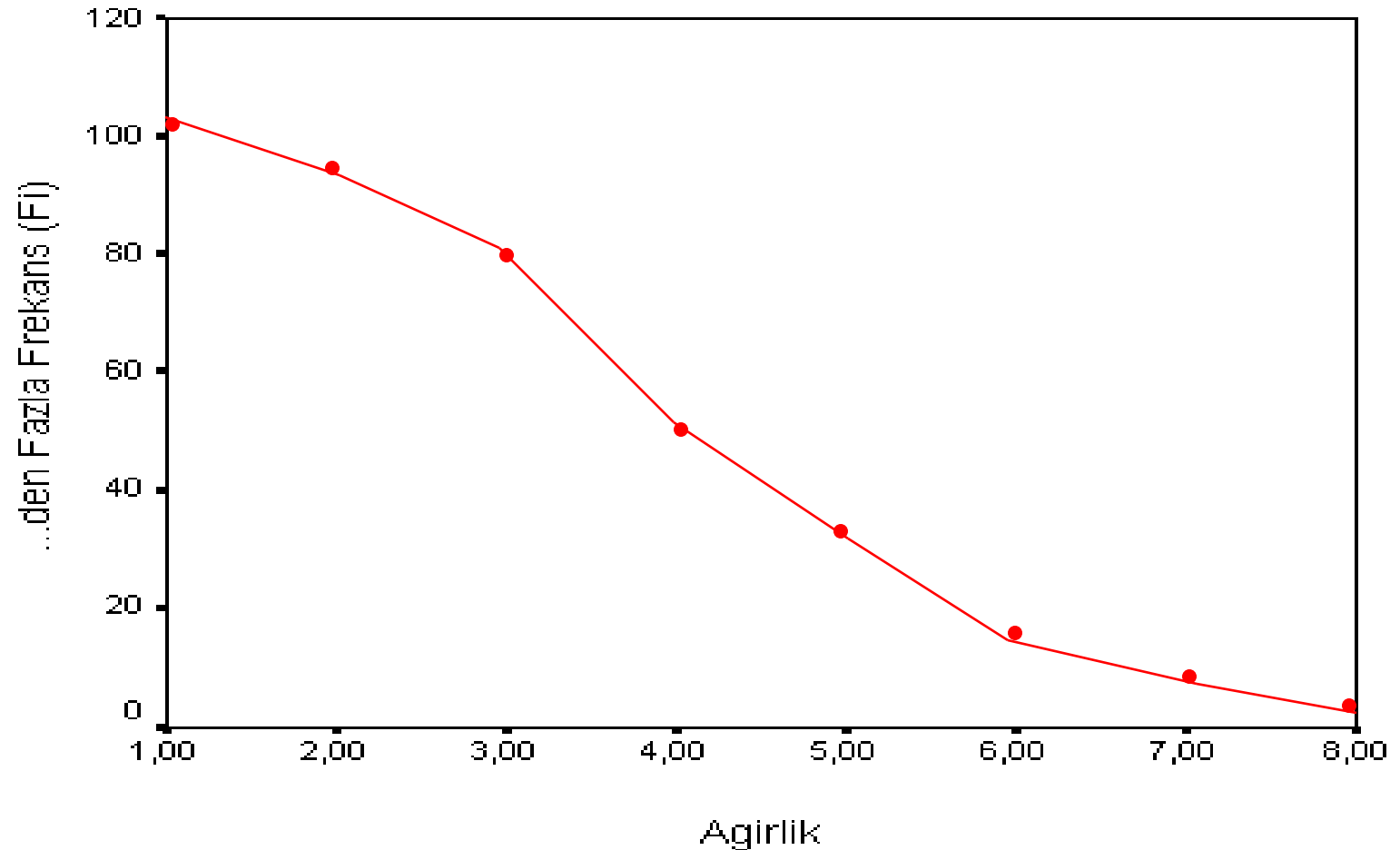


# EKLEMELİ SIKLIKLAR (..DEN FAZLA $F_i$ )

(1.45 DEN FAZLA AĞIRLIĞI OLAN 100 ÇOCUK VAR)

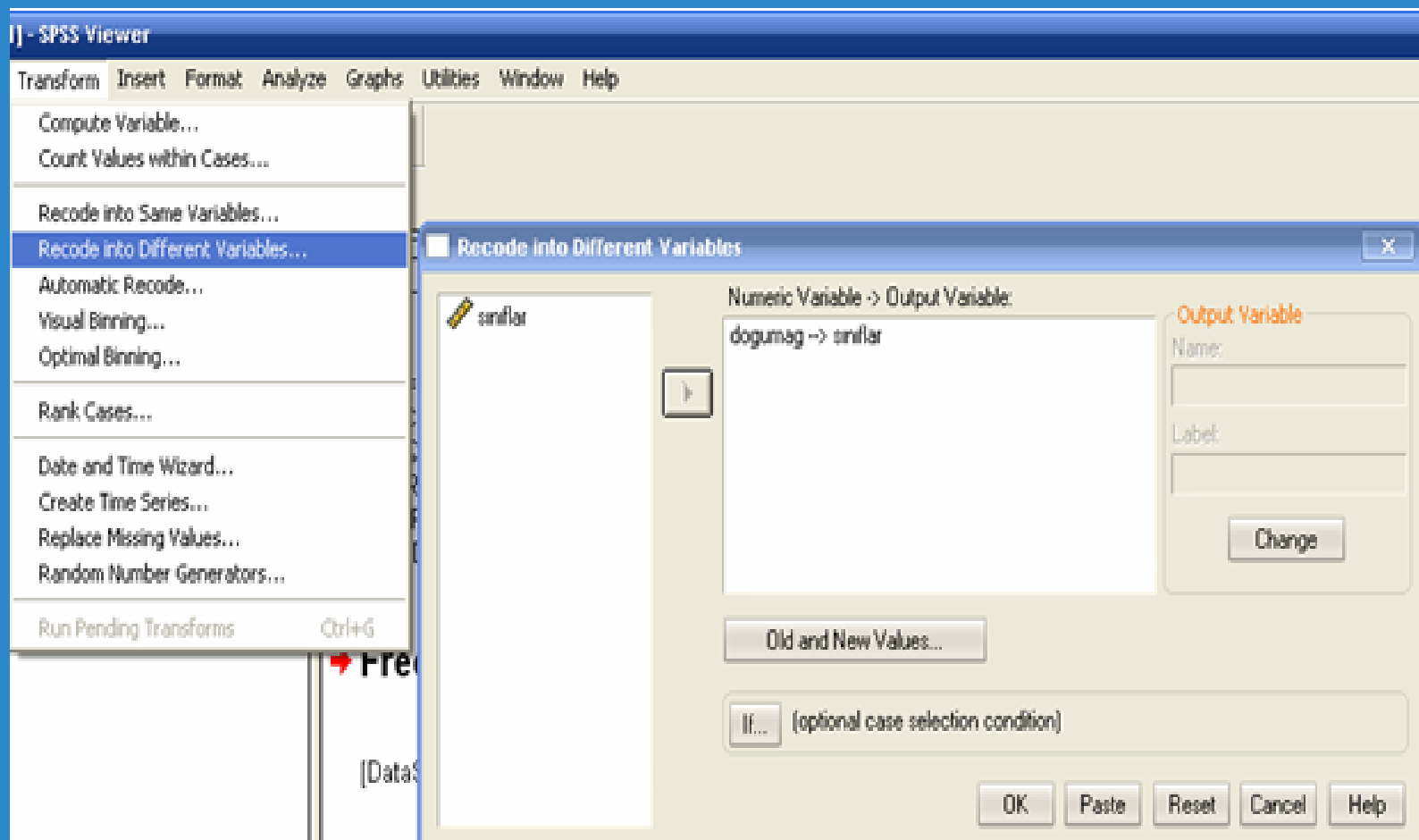
Sınıf sınırları	Sıklık (frekans= $f_i$ )	....den Fazla ( $F_i$ )
1.45 - 1.94	2	100
1.95 - 2,44	18	98
2,45 – 2,94	24	80
2,95 – 3,44	19	56
3,45 – 3,94	18	37
3,95 – 4,44	9	19
4,45 – 4,94	6	10
4,95 – 5,44	4	4
	100	

# EKLEMELİ SIKLIKLAR (..DEN FAZLA $F_i$ ) GRAFIĞİ





# SPSS İLE GRAFİK



# ELLE KODLAMA

Recode into Different Variables: Old and New Values

**Old Value**

☒ Value:

☐ System-missing

☐ System- or user-missing

☐ Range:

through

☐ Range, LOWEST through value:

☐ Range, value through HIGHEST:

☐ All other values

**New Value**

☒ Value:

☐ System-missing

☐ Copy old value(s)

Old --> New:

Add

Change

Remove

1.5 thru 1.9 --> 1  
2 thru 2.4 --> 2  
2.5 thru 2.9 --> 3  
3 thru 3.4 --> 4  
3.5 thru 3.9 --> 4  
4 thru 4.4 --> 5  
4.5 thru 4.9 --> 6  
5 thru 5.4 --> 7  
5.5 thru 5.9 --> 8

☐ Output variables are strings Width:

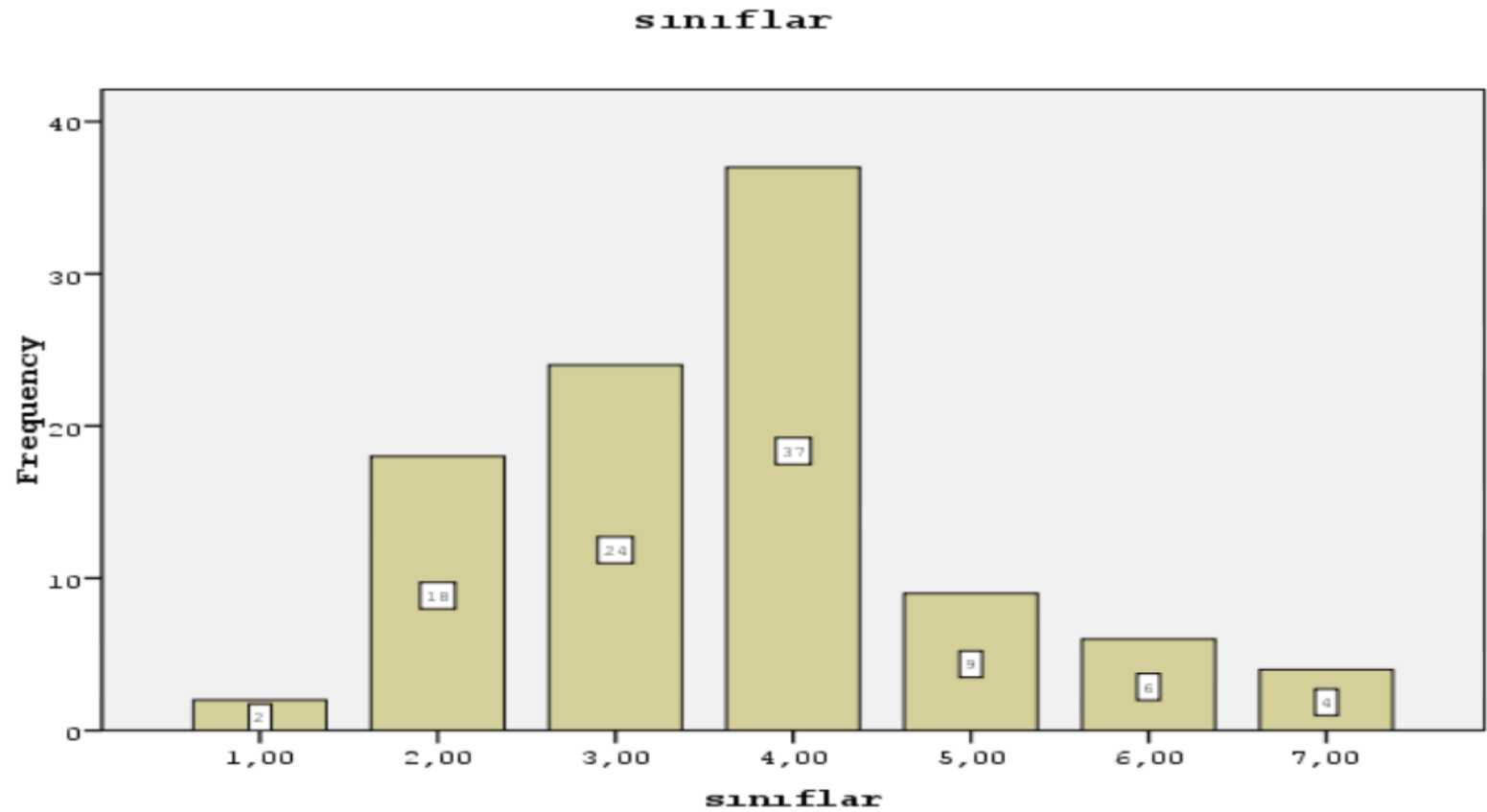
☐ Convert numeric strings to numbers ('5' -> 5)

Continue Cancel Help

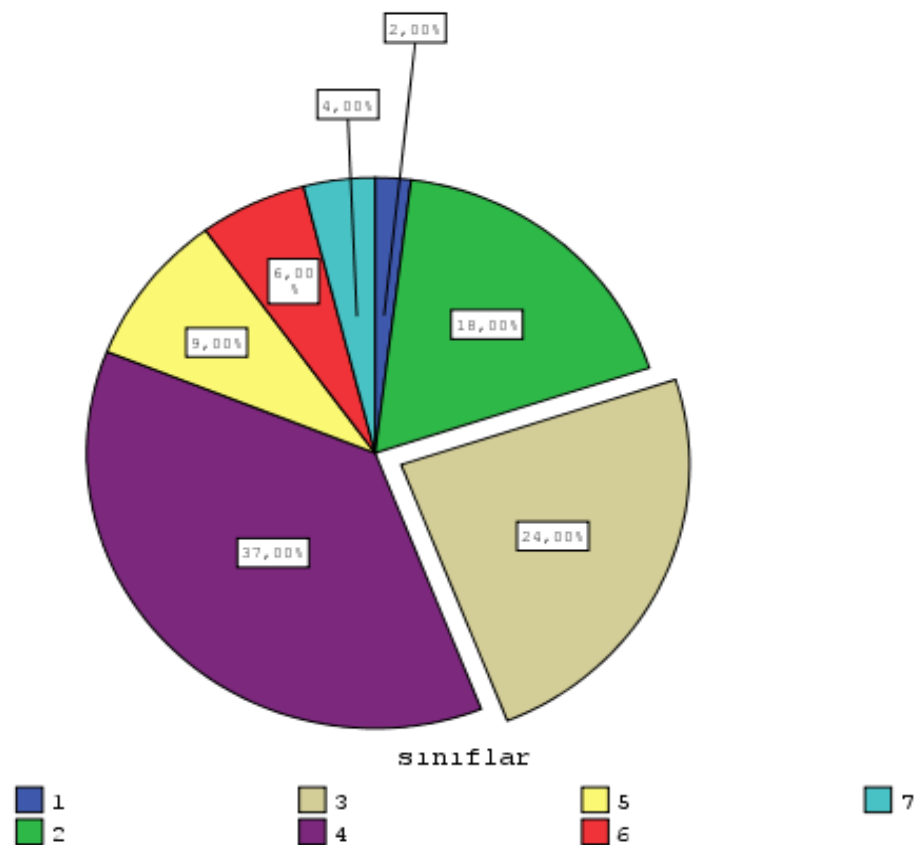
### sınıflar

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	2	2,0	2,0	2,0
	2,00	18	18,0	18,0	20,0
	3,00	24	24,0	24,0	44,0
	4,00	37	37,0	37,0	81,0
	5,00	9	9,0	9,0	90,0
	6,00	6	6,0	6,0	96,0
	7,00	4	4,0	4,0	100,0
Total		100	100,0	100,0	

# SÜTUN GRAFİK



# PASTA GRAFİK



# OTOMATİK SINIFLANDIRMA

cocuk agirlik.spo [Document 1] - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Output Log Frequencies Title Notes Active Dataset Statistics sýnýflar Bar Chart

**Make Cutpoints**

☒ Equal Width Intervals  
Intervals - fill in at least two fields

First Cutpoint Location: 1,90  
Number of Cutpoints: 7  
Width: 0,5  
Last Cutpoint Location: 4,90

☐ Equal Percentiles Based on Scanned Cases  
Intervals - fill in either field

Number of Cutpoints:   
Width(%):

☐ Cutpoints at Mean and Selected Standard Deviations Based on Scanned Cases

☐ +/- 1 Std. Deviation  
☐ +/- 2 Std. Deviation  
☐ +/- 3 Std. Deviation

Apply will replace the current cutpoint definitions with this specification. A final interval will include all remaining values: N cutpoints produce N+1 intervals.

**GGraph**

Visual Binning

Scanned Variable List: L. Variable  
çocuk doğum ağırlığı-kg [d]

Current Variable: dogumag  
Binned Variable: çocuk doğum ağırlığı-kg (Binned)

Minimum: 1,70 Nonmissing Values Maximum: 5,20

Enter interval cutpoints or click Make Cutpoints for automatic intervals. A cutpoint value of 10, for example, defines an interval starting above the previous interval and ending at 10.

Grid:

	Value	Label
1	1,90	
2	2,40	
3	2,90	
4	3,40	
5	3,90	
6	4,40	
7	4,90	
8	HIGH	
9		

Upper Endpoints  
☒ Included (<=)  
☐ Excluded (<)

Make Cutpoints...  
Make Labels

☐ Reverse scale

OK Paste Reset Cancel Help

sınıflar

3 4 5 6 7

SPSS Processor is ready H: 33,73 , W: 16,95

Intern... Bilgis... KOSE YÜKSE... kalitati... verida... poliso... Micros... \*cucuk... Synta... cocuk ... ZoomT... Belge1... 23:04

# TRANSFORM>VISUAL BINNING (OTOMATİK KODLAMA)

**Visual Binning**

Scanned Variable List:

L.	Variable
1	çocuk doğum ağırlığı-kg [d

Current Variable: dogumag

Binned Variable:

Minimum: 1,70 Nonmissing Values Maximum: 5,20

Label: çocuk doğum ağırlığı-kg

çocuk doğum ağırlığı-kg (Binned)

Enter interval cutpoints or click Make Cutpoints for automatic intervals. A cutpoint value of 10, for example, defines an interval starting above the previous interval and ending at 10.

Grid:

	Value	Label
1	1,90	<= 1,90
2	2,40	1,91 - 2,40
3	2,90	2,41 - 2,90
4	3,40	2,91 - 3,40
5	3,90	3,41 - 3,90
6	4,40	3,91 - 4,40
7	4,90	4,41 - 4,90
8	HIGH	4,91+
9		

Cases Scanned: 100

Missing Values: 0

Copy Bins

From Another Variable...

To Other Variables...

Upper Endpoints

☒ Included (<=)

☐ Excluded (<)

Make Cutpoints...

Make Labels

☐ Reverse scale

OK Paste Reset Cancel Help

Yeni isim ver

Tanım yapılır

Label oluşturur