



HACETTEPE ÜNİVERSİTESİ  
İSTATİSTİK BÖLÜMÜ

## İST155 İSTATİSTİĞE GİRİŞ I

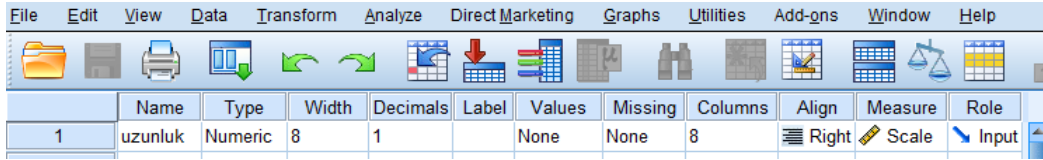
### UYGULAMA 3

Ar. Gör. Dr. Derya Turfan – Ar. Gör. Dr. Murat Arat

1) 50 bitkinin yaprak uzunlukları aşağıda verilmiştir. Verilerin sıklık çizelgesini düzenleyiniz.

13,8	16,8	18,6	17,4	17,5	16,5	15,9	12,8	14,6	14,9
17,0	15,2	13,6	12,3	13,4	12,7	18,5	17,7	10,6	20,3
12,2	10,7	13,8	18,6	11,7	15,7	21,2	16,3	15,3	11,7
11,9	17,8	19,4	14,3	18,1	14,4	11,1	13,3	12,4	15,6
20,1	17,7	11,6	11,4	9,6	15,6	11,2	17,4	9,7	20,1

İlk olarak, *Variable View* penceresinde yaprak uzunluklarına ait “uzunluk” değişkeni tanımlanarak *Data View* penceresinde bu değişkene ait gözlem değerleri girilir.

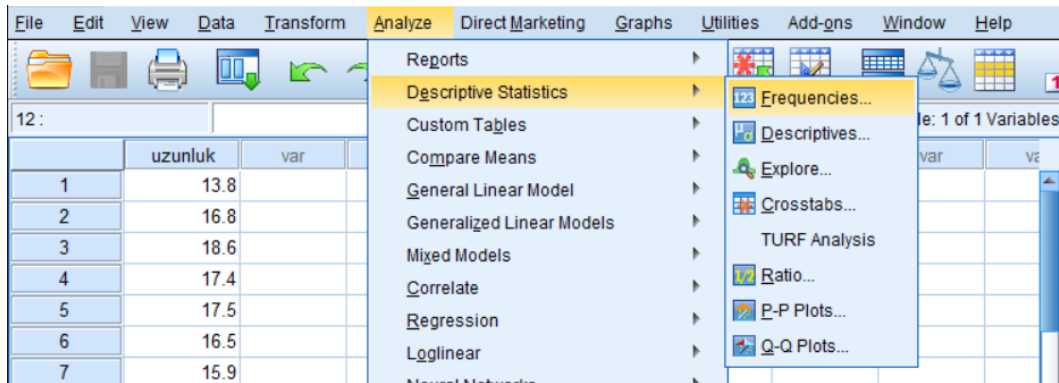


Data View kısmında ise 50 gözlem biriminin girişi tamamlanır. İlk 5 verinin girişi örnek olarak verilmiştir.

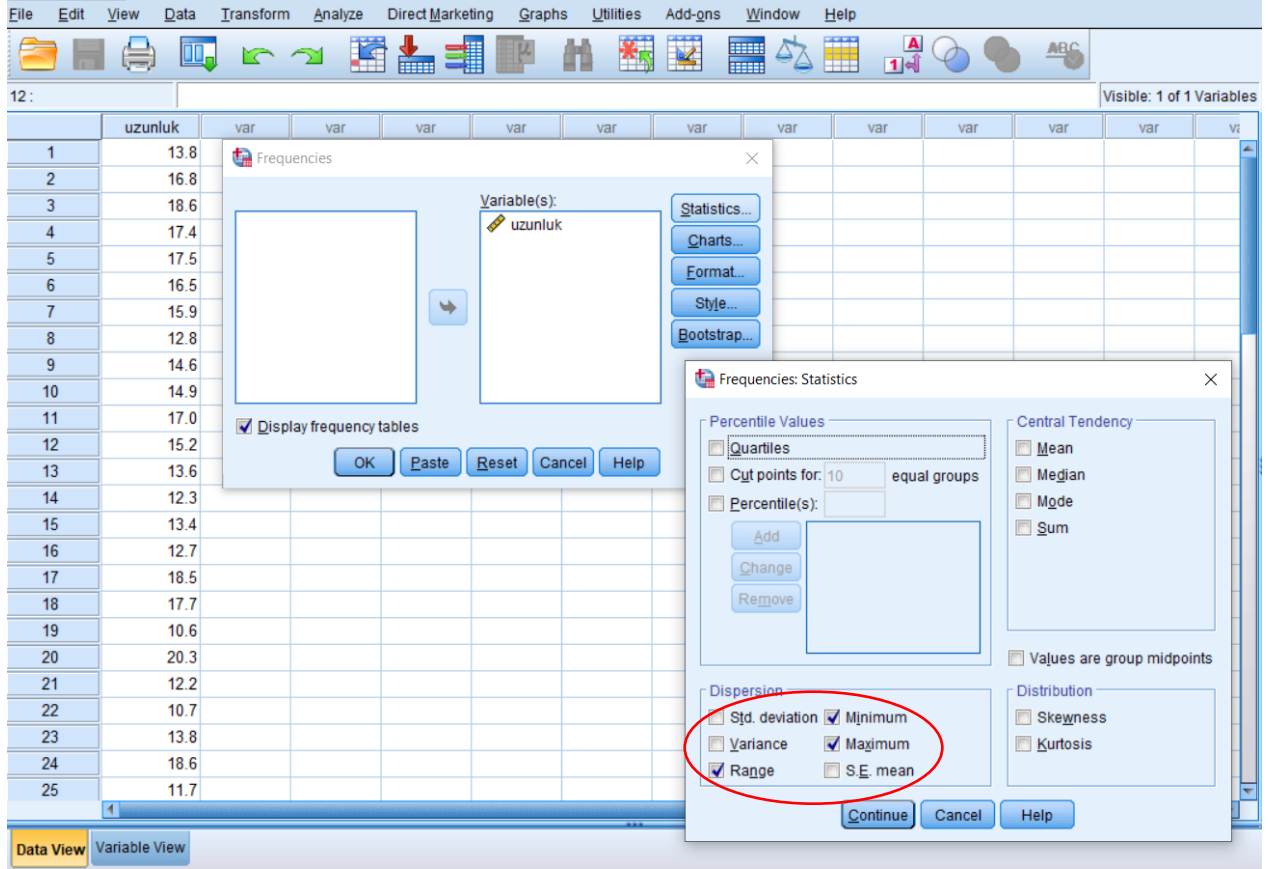


	uzunluk	var
1	13.8	
2	16.8	
3	18.6	
4	17.4	
5	17.5	
...		

Sıklık çizelgesinin düzenlenmesinde ilk adım olarak veri setinde yer alan minimum, maksimum değerler ile birlikte dağılım genişliği elde edilir.



*Frequencies* seçeneği tıklandıktan sonra açılan pencerede sol kutuda yer alan değişken(ler) görülmektedir. Sorumuz için sadece tek bir değişken yer almaktadır ve sadece bu değişkene ait işlemler yapılacağı için “uzunluk” değişkeni seçilerek sağ kutu (Variable) kısmına gönderilir. *Statistics* kısmı tıklanarak veri seti için minimum, maksimum ve dağılım genişliği (range) elde edilir.



Statistics		
uzunluk		
N	Valid	50
	Missing	0
Range		11.6
Minimum		9.6
Maximum		21.2



Çıktı sayfasında görülen bu değerlerden yararlanılarak sınıf sayısı ( $k$ ) ve sınıf aralığı ( $c$ ) elde edilir.

$$\text{sınıf sayısı } (k) = 1 + 3,3 \log(n) = 1 + 3,3 \log(50) \cong 6,61 \cong 7$$

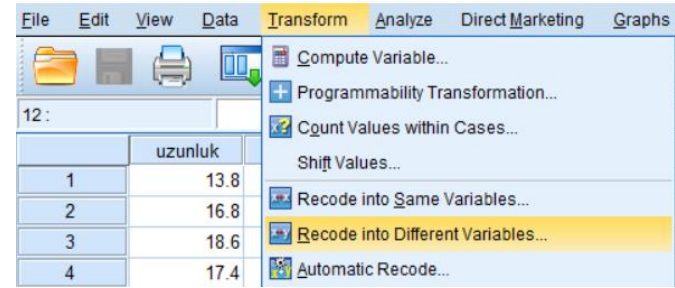
$$\text{sınıf aralığı } (c) = \frac{\text{Dağılım genişliği} + \text{son haneye 1 ekle}}{k} = \frac{11,6 + 0,1}{7} \cong 1,67 \cong 1,7$$

Belirlenen sınıf sayısı ve sınıf aralığı ile sınıflara ait sıklıkların elde edilebilmesi için sınıfların alt sınır ve üst sınır değerleri elde edilir.

Sınıf	Alt Sınır	Üst Sınır
1	9,6	11,2
2	11,3	12,9
3	13,0	14,6
4	14,7	16,3
5	16,4	18,0
6	18,1	19,7
7	19,8	21,4

Sınıflar arası geçişte her sınıf aralığının 1,7 olduğuna dikkat edilmelidir.

Artık sınıf bazında konuşacağımız için SPSS programında nicel değerlerimiz nitel değişkene yani sınıflara dönüştürülür. Bu işlem için aşağıda yer alan adımlar takip edilmelidir.



The first screenshot shows the SPSS 'Transform' menu with 'Recode into Different Variables...' selected. The second screenshot shows the 'Recode into Different Variables' dialog box. The 'Numeric Variable -> Output Variable:' section is highlighted with a red circle, showing 'uzunluk -> uzunluk\_sınıf'. The 'Output Variable' section is also highlighted with a red circle, showing 'Name: uzunluk\_sınıf'. A red arrow points to the 'If...' button at the bottom of the dialog box.

File Edit View Data Transform Analyze Direct Marketing Graphs

12 :  
uzunluk  
1 13.8  
2 16.8  
3 18.6  
4 17.4

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

26 :  
uzunluk var var var var var var var var var  
1 13.8  
2 16.8  
3 18.6  
4 17.4  
5 17.5  
6 16.5  
7 15.9  
8 12.8  
9 14.6  
10 14.9  
11 17.0  
12 15.2  
13 13.6  
14 12.3  
15 13.4  
16 12.7  
17 18.5

Recode into Different Variables

Numeric Variable -> Output Variable:  
uzunluk -> uzunluk\_sınıf

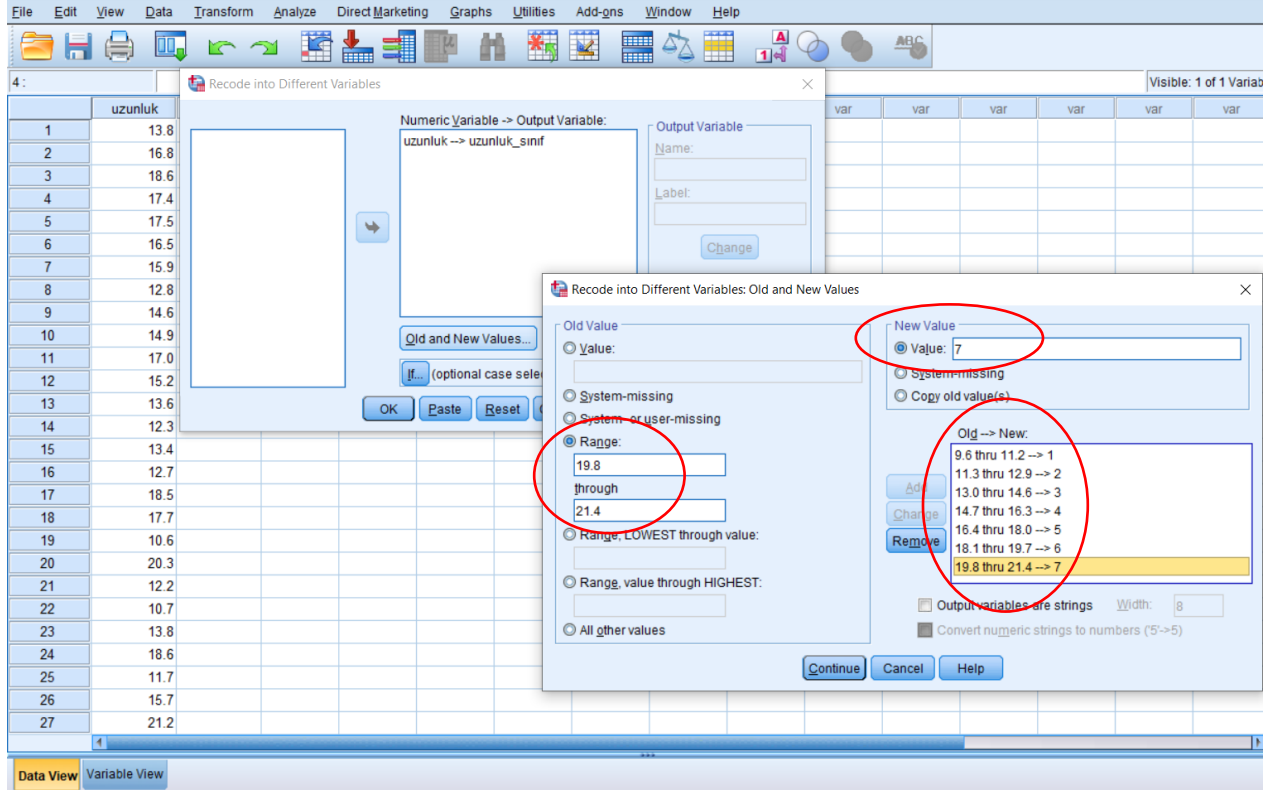
Output Variable  
Name:  
uzunluk\_sınıf  
Label:  
Change

Old and New Values...

If... (optional case selection condition)

OK Paste Reset Cancel Help

“uzunluk” değişkeninden yararlanarak yeni bir değişken elde ettiğimiz için bu aşamada “uzunluk\_sınıf” adında yeni bir değişken tanımlarız. Bu değişken, ham verilerden elde edilen sınıflardan oluşacağı için değerlerin girilmesinde ilk olarak *Old and New Values...* seçeneği tıklanır ve açılan sayfada ise *Range* kısmından yararlanılarak her bir sınıfın alt ve üst değerleri girilir ve belirlenen 7 sınıfın tanımlaması yapılır.

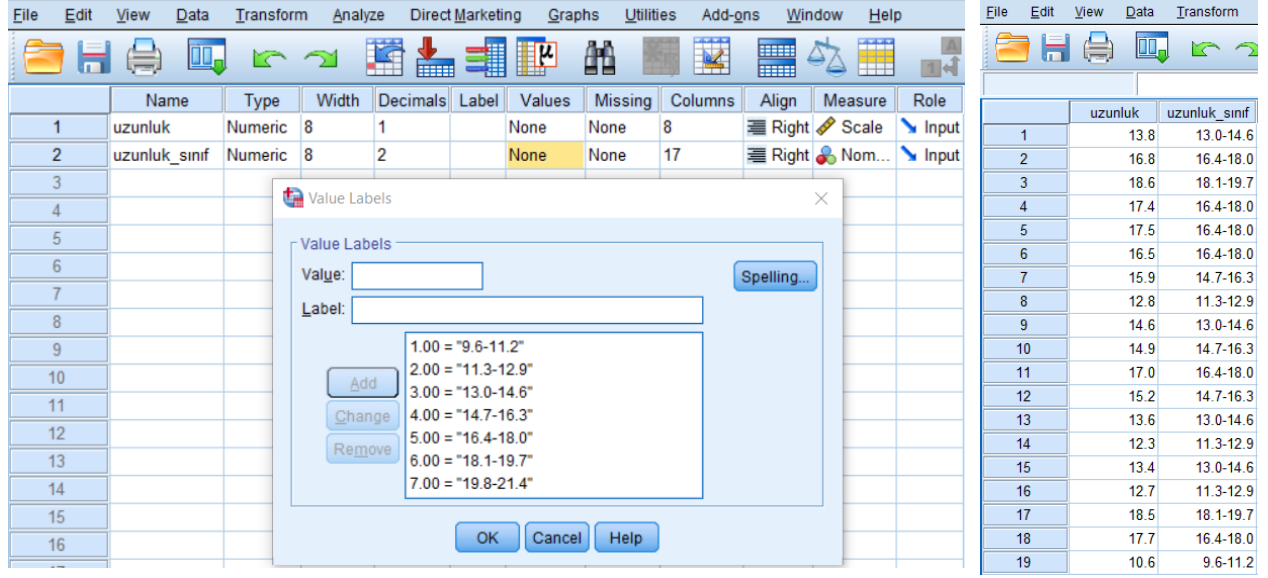


*Data View* penceresinde oluşturulan yeni değişken görülür.

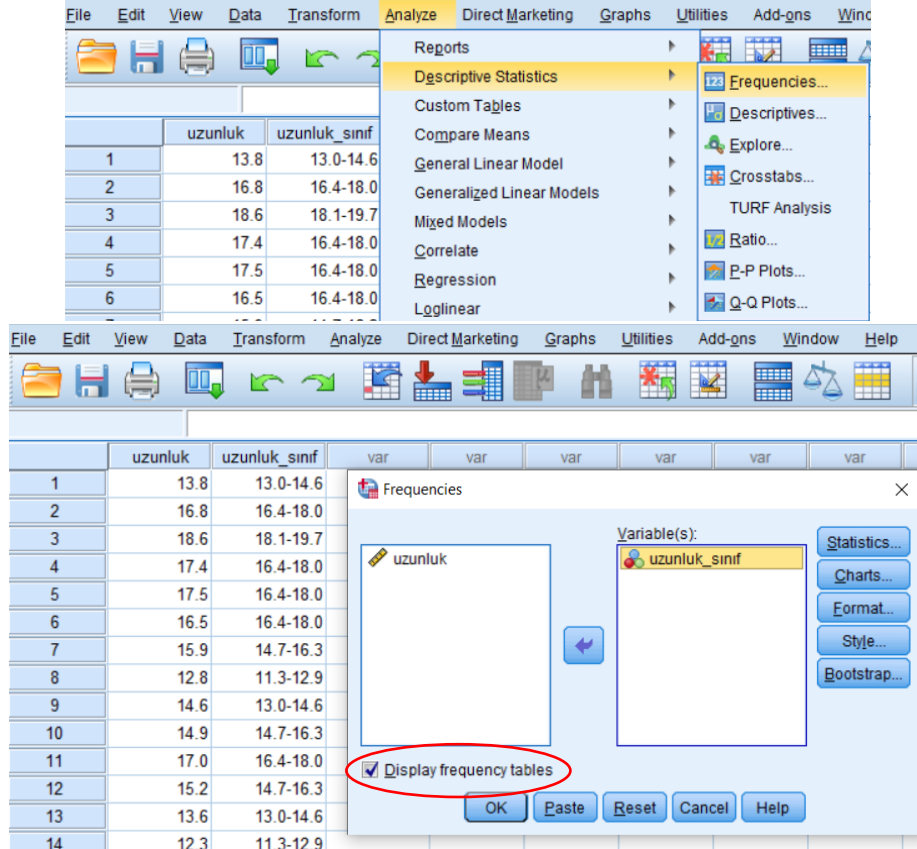


	uzunluk	uzunluk_sınıf
1	13.8	3.00
2	16.8	5.00
3	18.6	6.00
4	17.4	5.00
5	17.5	5.00
6	16.5	5.00
7	15.9	4.00
8	12.8	2.00
9	14.6	3.00
10	14.9	4.00
11	17.0	5.00
12	15.2	4.00
13	13.6	3.00
14	12.3	2.00
15	13.4	3.00

“uzunluk\_sınıf” değişkeni için belirlenen sınıf değerlerine göre etiketleme yapılır.



Sonuç olarak, elde edilen tüm bilgilerle sıklık çizelgesi tamamlanır. Bu işlem için *Frequencies* seçeneği seçilir. Açılan pencerede artık “uzunluk\_sınıf” değişkeni üzerinden işlem yapacağımız için bu değişken seçilir ve sayfada yer alan *Display frequency tables* seçeneği işaretlenerek işlem tamamlanır.



Çıktı sayfasında aşağıda yer alan tablo elde edilir. SPSS programında sıklık çizelgesi için sıklık(frekans), göreceli sıklık, birikimli göreceli sıklıklar elde edilebilmektedir.

uzunluk sınıf		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	9.6-11.2	6	12.0	12.0	12.0
	11.3-12.9	10	20.0	20.0	32.0
	13.0-14.6	8	16.0	16.0	48.0
	14.7-16.3	8	16.0	16.0	64.0
	16.4-18.0	9	18.0	18.0	82.0
	18.1-19.7	5	10.0	10.0	92.0
	19.8-21.4	4	8.0	8.0	100.0
	Total	50	100.0	100.0	

Sınıf	Alt Sınır	Üst Sınır	$f_i$	$p_i$	$S_i$	$F_i$	$\%F_i$
1	9,6	11,2	6	0,12	10,4	6	0,12
2	11,3	12,9	10	0,20	12,1	16	0,32
3	13,0	14,6	8	0,16	13,8	24	0,48
4	14,7	16,3	8	0,16	15,5	32	0,64
5	16,4	18,0	9	0,18	17,2	41	0,82
6	18,1	19,7	5	0,10	18,9	46	0,92
7	19,8	21,4	4	0,08	20,6	<b>50</b>	<b>1</b>

- ✓ Yaprak uzunluğu 13 ile 14,6 arasında olan bitki sayısı nedir?
- ✓ Yaprak uzunluğu 16,4 ile 18 arasında olan bitkilerin toplam bitkiler içindeki oranı nedir?
- ✓ Yaprak uzunluğu 14,6 ve az olan bitki sayısı ve oranı nedir?
- ✓ Yaprak uzunluğu 19,8 ve fazla olan bitki sayısı ve oranı nedir?

2) Bir sigorta şirketi müşterilerine daha iyi hizmet sunabilmek için şirket politikasında yenilikler yapmak istemektedir. Müşteri memnuniyetini sağlayabilmek için yolda kalan müşterilerin mağduriyetini en kısa zamanda gidermek istemektedir. Bunun için bir ay boyunca yolda kalan sürücülerin yardım çağrısı sayıları kaydedilmiştir. Verilerin sıklık çizelgesinin düzenleyiniz.

28	122	217	130	120	86	80	90	120	140	81	70	40	145	187	113
90	68	174	194	170	100	75	104	97	75	123	100	82	109	120	

Birinci soruda yer alan adımlar ikinci soru için de geçerlidir. İlk olarak “cagri” adlı değişken *Variable View* sayfasında tanımlanır ve gözlem birimleri *Data View* sayfasında girilir.

File	Edit	View	Data	Transform	Analyze	Direct Marketing	Graphs	Utilities	Add-ons	Window	Help
1	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	cagri	Numeric	8	0		None	None	8	Right	Scale	Input

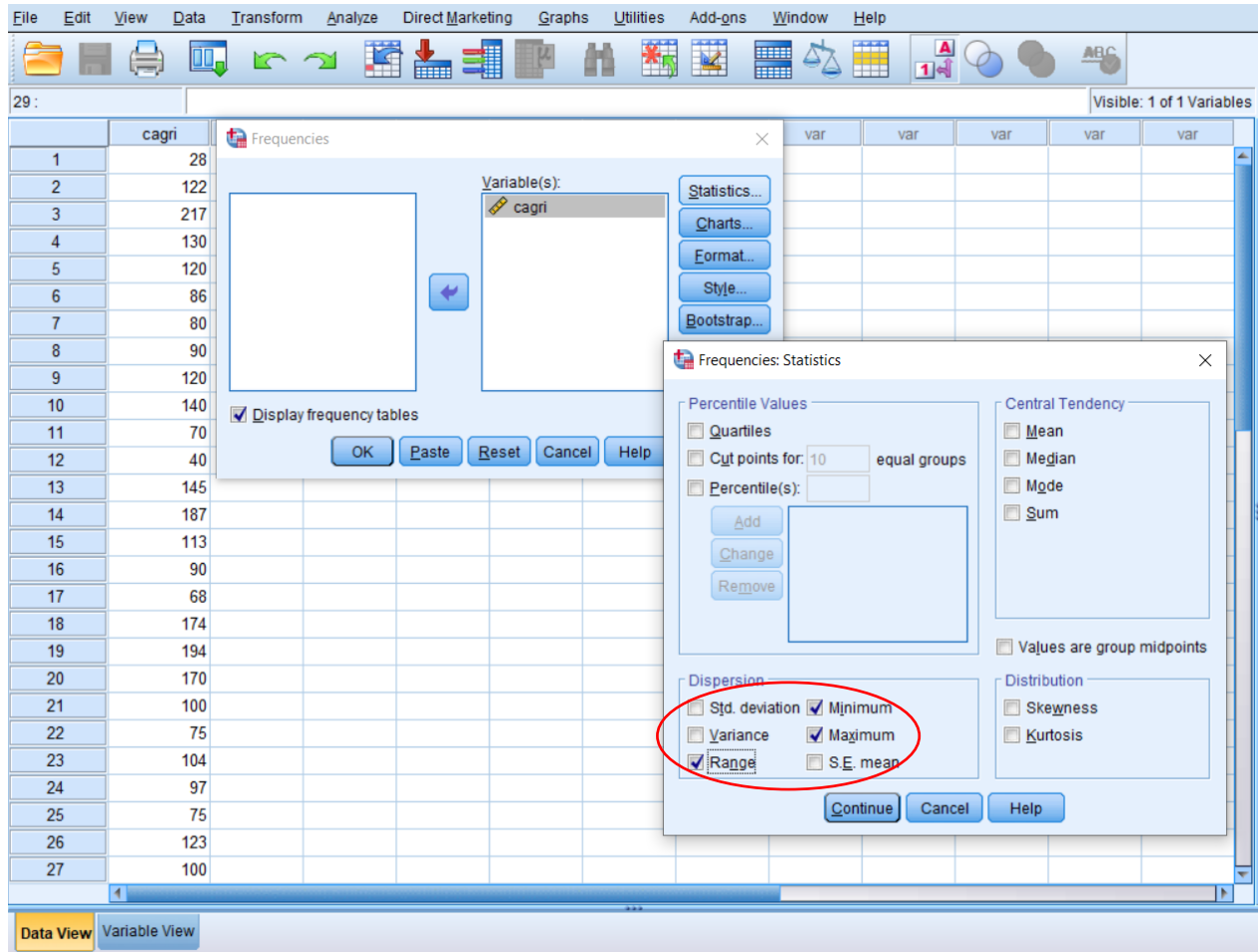
  

File	Edit	View	Data
8 :			
1			28
2			122
3			217
4			130

File	Edit	View	Data
29 :			
28			82
29			109
30			120
31			81

*Analyze* → *Descriptive Statistics* → *Frequencies* adımları izlenerek çağrılar arasında minimum değer, maksimum değer ve dağılım genişliği elde edilir.



Statistics		
cagri		
N	Valid	31
	Missing	0
Range		189
Minimum		28
Maximum		217



$$\text{sınıf sayısı } (k) = 1 + 3,3 \log(31) \cong 5,92 \cong 6$$

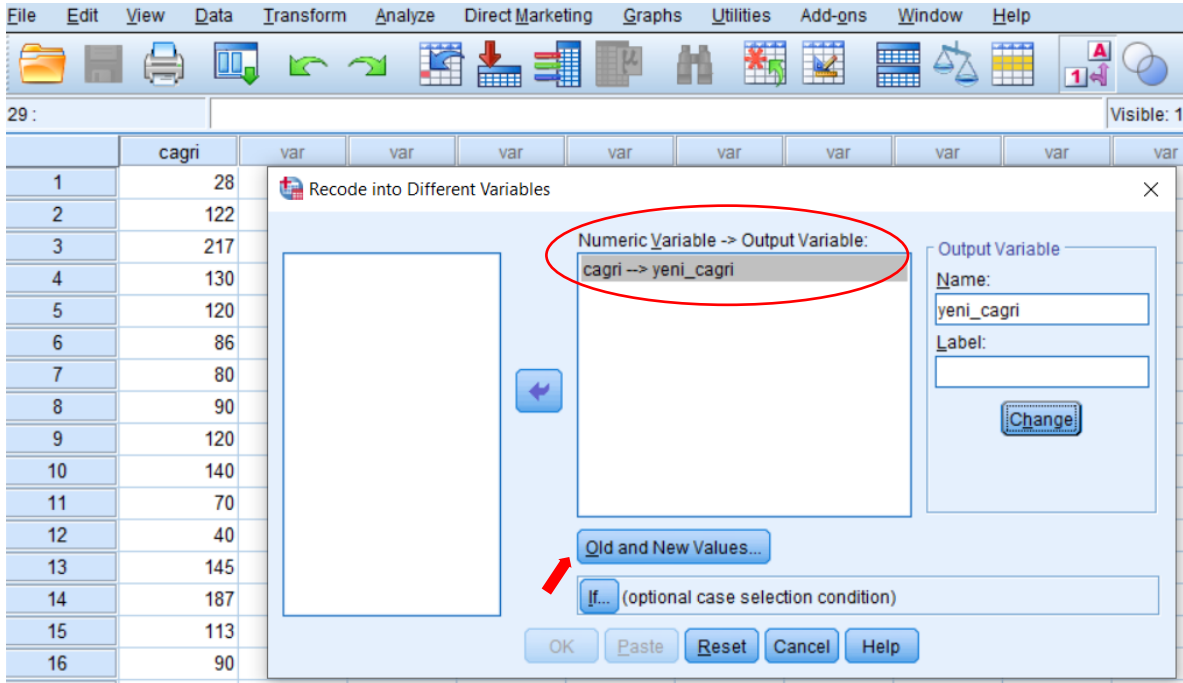
$$\text{sınıf aralığı } (c) = \frac{189 + 1}{6} = 31,67 \cong 32$$

Belirlenen sınıf sayısı ve sınıf aralığı ile sınıflara ait sıklıkların elde edilebilmesi için sınıfların alt sınır ve üst sınır değerleri elde edilir:

Sınıf	Alt Sınır	Üst Sınır
1	28	59
2	60	91
3	92	123
4	124	155
5	156	187
6	188	219

Sınıflar arası geçişte her sınıf aralığının 32 olduğuna dikkat edilmelidir...

Çağrı sayılarını gösteren nicel değişkenimiz nitel değişkene yani sınıflara dönüştürülür. Bu işlem **Transform → Recode into Different Variables** sekmesine tıklanarak aşağıda yer alan adımlar takip edilmelidir. “cagri” değişkeninden yararlanarak yeni bir değişken elde ettiğimiz için bu aşamada “yeni\_cagri” adında yeni bir değişken tanımlarız. Bu değişken, ham verilerden elde edilen sınıflardan oluşacağı için değerlerin girilmesinde ilk olarak Old and New Values... seçeneği tıklanır.



Açılan sayfada ise *Range* kısmından yararlanılarak her bir sınıfın alt ve üst değerleri girilir ve belirlenen 6 sınıf elde edilir. Daha sonra yeni\_cagri değişkeni için belirlenen sınıf değerlerine göre etiketleme yapılır.



Old Value

☐ Value:

☐ System-missing

☐ System- or user-missing

☒ Range:

188 through 219

☐ Range, LOWEST through value:

☐ Range, value through HIGHEST:

☐ All other values

New Value

☒ Value: 6

☐ System-missing

☐ Copy old value(s)

Old --> New:

28 thru 59 --> 1

60 thru 91 --> 2

92 thru 123 --> 3

124 thru 155 --> 4

156 thru 187 --> 5

188 thru 219 --> 6

Output variables are strings Width: 8

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

Continue Cancel Help

File Edit View Data Transform Analyze Dir

10 : cagri 140

	cagri	yeni_cagri
1	28	1.00
2	122	3.00
3	217	6.00
4	130	4.00
5	120	3.00



File Edit View Data Transform Analyze Dir

	cagri	yeni_cagri
1	28	28-59
2	122	92-123
3	217	188-219
4	130	124-155
5	120	92-123

**Analyze → Descriptive Statistics → Frequencies** adımları izlenerek “yeni\_cagri” değişkeni üzerinden sıklık çizelgesi elde edilir.

**yeni\_cagri**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	28-59	2	6.5	6.5	6.5
	60-91	10	32.3	32.3	38.7
	92-123	11	35.5	35.5	74.2
	124-155	3	9.7	9.7	83.9
	156-187	3	9.7	9.7	93.5
	188-219	2	6.5	6.5	100.0
	Total	31	100.0	100.0	

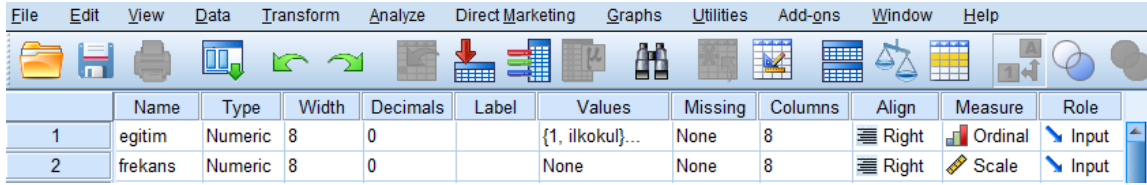
Sınıf	Alt Sınır	Üst Sınır	$f_i$	$p_i$	$S_i$	$F_i$	$\%F_i$
1	28	59	2	0,065	43,5	2	0,065
2	60	91	10	0,323	75,5	12	0,387
3	92	123	11	0,355	107,5	23	0,742
4	124	155	3	0,097	139,5	26	0,839
5	156	187	3	0,097	171,5	29	0,935
6	188	219	2	0,065	203,5	31	1

- ✓ Günde 60 ile 91 arasında kaç çağrı yapılmıştır?
- ✓ Çağrı sayısının 59 ve daha az olmasının yüzdesi nedir?
- ✓ Çağrı sayısının 123 ve daha az olmasının yüzdesi nedir?
- ✓ Çağrı sayısının 156 ve daha çok olmasının yüzdesi nedir?

3) Bir işyerinde sigara için 500 kişinin eğitim düzeylerine göre dağılımları verilmiştir. Sıklık çizelgesini düzenleyiniz.

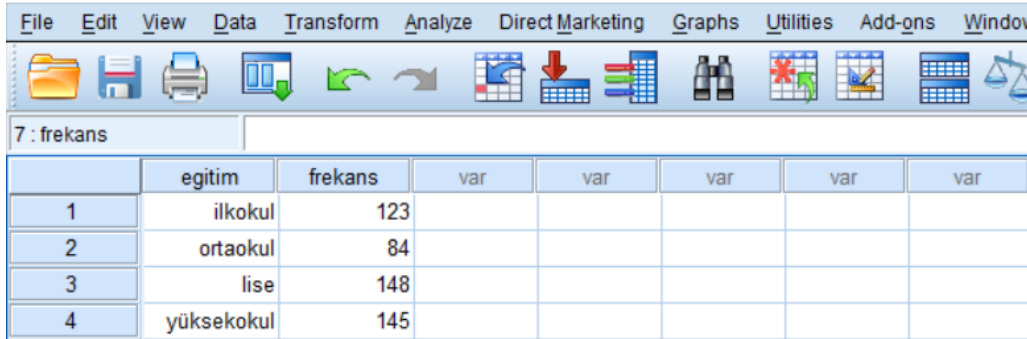
Eğitim Düzeyi	f <sub>i</sub>
İlkokul	123
Ortaokul	84
Lise	148
Yüksekokul	145

Bu soruda sınıflarımız eğitim düzeyi olup 123 kişinin ilkokul, 84 kişinin ortaokul, 148 kişinin lise ve kalan 145 kişinin ise yüksekokul mezunu olduğu belirtilmiştir. Programda bu soru için ilkokul 1, ortaokul 2, lise 3 ve yüksekokul 4 olarak etiketlenerek belirtilen sayılarda 500 kişinin eğitim durumunun girişi yapılabilir. Fakat bu zaman ve emek açısından kayıp olacağından ağırlıklandırma yapılır. Bu işlem için ilk olarak *Variable View* penceresinde eğitim ve bu eğitim seviyelerinde kaç kişi olduğunu belirtmek adına frekans değişkenleri tanımlanır. Eğitim düzeyleri için de etiketleme işlemi yapılır.



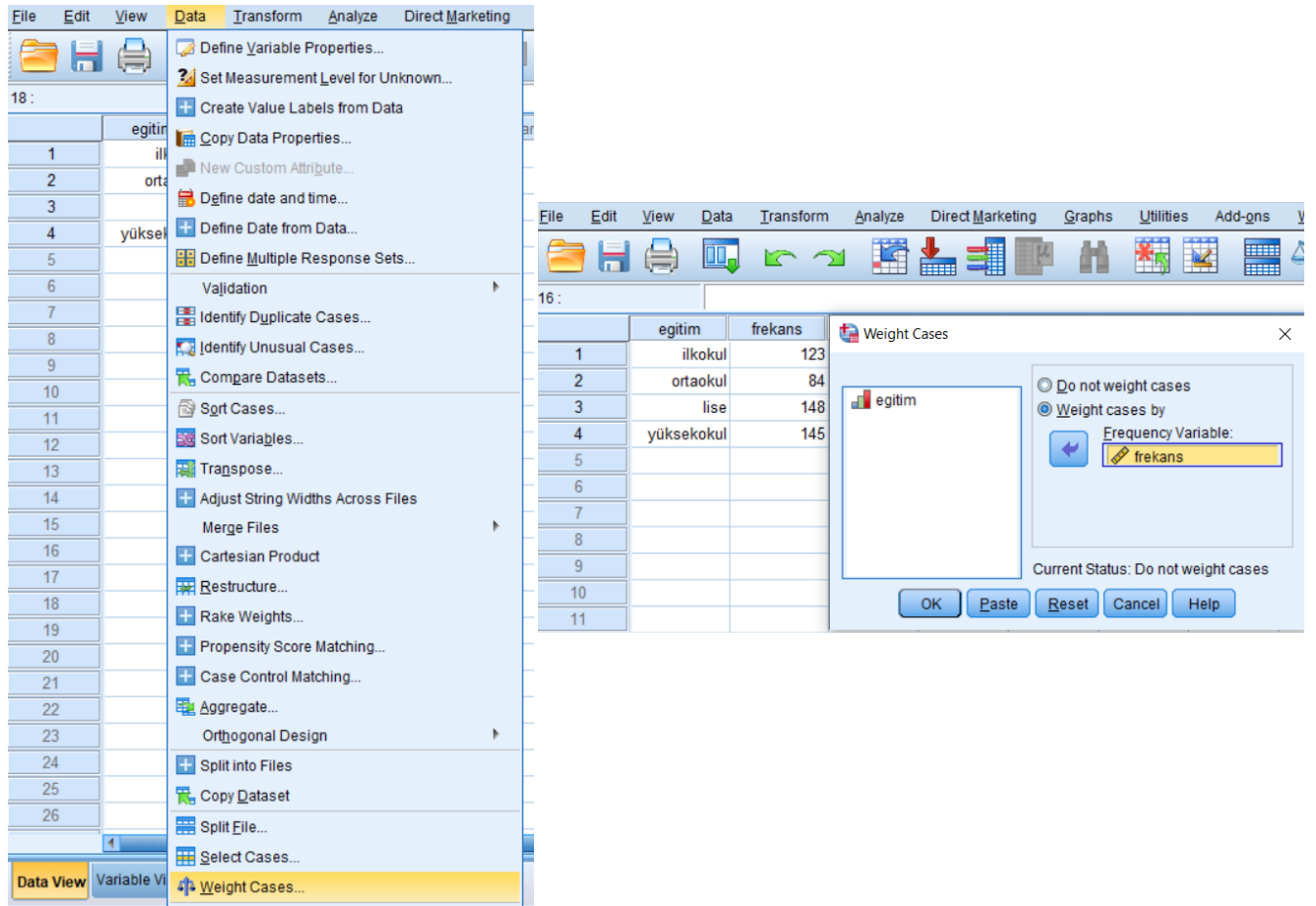
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	egitim	Numeric	8	0		{1, ilkokul}...	None	8	Right	Ordinal	Input
2	frekans	Numeric	8	0		None	None	8	Right	Scale	Input

Bu değişkenlere bağlı olarak *Data View* sayfasında ise verilerin girişleri yapılır.



7 : frekans							
	egitim	frekans	var	var	var	var	var
1	ilkokul	123					
2	ortaokul	84					
3	lise	148					
4	yuksekokul	145					

SPSS programına 123 kişinin ilkokul, 84 kişinin ortaokul, 148 kişinin lise ve 145 kişinin ise yüksekokul eğitim düzeyine sahip olduğunu tanımlatmak için **Data→Weight Cases** adımları izlenir.



Bu işlemle, “**frekans**” değişkenine göre ağırlıklandırma yapılır. Bu işlemin aktif olup olmadığı *Data View-Variable View* sayfalarında görülebilen sağ alt kısımdan kontrol edilmektedir. Bu kısımda *Weight On* yazıyor ise bu ağırlıklandırmanın var olduğu anlamına gelmektedir. İşlem bittikten sonra yine *Weight Cases* kısmında **Reset** tuşu tıklanarak, ağırlıklandırma işlemi aktif olmaktan çıkarılır.

Ağırlıklandırmanın aktif olduğu kontrol edilerek sıklık çizelgesi için *Analyze → Descriptive Statistics → Frequencies* adımları izlenerek “*egitim*” değişkeni üzerinden sıklık çizelgesi elde edilir.

		egitim			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ilkokul	123	24.6	24.6	24.6
	ortaokul	84	16.8	16.8	41.4
	lise	148	29.6	29.6	71.0
	yüksekokul	145	29.0	29.0	100.0
	Total	500	100.0	100.0	

- ✓ İlkokul mezunu olan kaç kişi sigara içmektedir?
- ✓ Sigara içen 500 kişinin içinde yüksekokul mezunu olanların oranı nedir?
- ✓ Sigara içenler içerisinde lise ve daha yüksek düzeyde eğitim düzeyi olanların sayısı ve oranı nedir?

Sınıf	$f_i$	$p_i$	$F_i$	$\%F_i$
ilkokul	123	0,246	123	0,246
ortaokul	84	0,168	207	0,414
lise	148	0,296	355	0,71
yüksekokul	145	0,29	<b>500</b>	<b>1</b>

4) Hacettepe Üniversitesi Beytepe Sağlık merkezinde 2006 yılı Ocak ayında sağlık kontrolü yapılan kadın personelin konulan hastalık tanılarına göre dağılımı verilmiştir. Sıklık çizelgesini düzenleyiniz.

Hastalık	Tanı Sayısı
Hipertansiyon	12
Tiroit Hastalığı	9
Diyabet	3
Obezite	4
Osteoporoz	21
Hiperlipidemi	51

Bu soruda sınıflarımız hastalık türleri olup 12 kişinin hipertansiyon, 9 kişinin tiroit, ..., 51 kişinin ise hiperlipidemi hastalığına sahip olduğu görülmektedir. 3. sorudaki benzer mantıkla bu soru için de ağırlıklandırma yapılır. (1 → hipertansiyon, 2 → tiroit, ..., 6 → hiperlipidemi)

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	hastalık	Numeric	8	0		{1, hipertansiyon}...	None	8	Right	Nominal	Input
2	frekans	Numeric	8	0		None	None	8	Right	Scale	Input

	hastalık	frekans
1	hipertansiyon	12
2	tiroit	9
3	diyabet	3
4	obezite	4
5	osteoporoz	21
6	hiperlipidemi	51

Benzer olarak, “*frekans*” değişkenine göre **Data→Weight Cases** adımları izlenerek ağırlıklandırma yapılır. Ağırlıklandırmanın aktif olduğu kontrol edilerek sıklık çizelgesi için **Analyze→Descriptive Statistics→Frequencies** adımları ile “*hastalık*” değişkeni üzerinden sıklık çizelgesi elde edilir.

hastalık		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	hipertansiyon	12	12.0	12.0	12.0
	tiroit	9	9.0	9.0	21.0
	diyabet	3	3.0	3.0	24.0
	obezite	4	4.0	4.0	28.0
	osteoporoz	21	21.0	21.0	49.0
	hiperlipidemi	51	51.0	51.0	100.0
	Total	100	100.0	100.0	

- ✓ Kaç hastaya hipertansiyon tanısı konmuştur?
- ✓ Hastalık tanısı konulan hastaların içinde obezite tanısı konulanların oranı nedir?
- ✓ Hiperlipidemi tanısı konulanların sayısı ve oranı nedir?