# 

***Dördüncü Bölüm***

***Temel istatistikler***

******

***Hedefler***

Bu üniteyi çalıştıktan sonra;

* Verileri özetlemeyi öğrenecek
* Lokasyon ve Dağılım ölçülerini öğrenecek
* Verileri sunmak konusunda bilgi sahibi olacaksınız

## *Anahtar Kavramlar*

Nominal ölçek

Ordinal ölçek

Scale ölçek

Mod

Medyan

Standart Sapma

Ortalama

Histogram

Bar Chart

***İçindekiler***

1. Verileri özetlemek ve sunmak
2. Verileri özetlemek
   1. Lokasyon Ölçüleri
   2. Dağılım Ölçüleri
   3. Verileri Sunmak
3. **Verileri Özetlemek ve Sunmak**

Herhangi bir istatistiksel analizde ilk aşama verileri özetlemek ve sunmaktır. Basit bir özet tablosu veriler hakkında bize önemli bilgiler verebilir. Bunun yanı hatalar ve kayıp değerler hakkında da bilgi sahibi olabiliriz. Herhangi bir istatisitiksel teste başlamadan önce veri tablomuzun temiz olması önemlidir.

Değişik sayısal türde veriler için, değişik istatiksel analizler mevcuttur. İstatistiksel analizler için ne tür veriye sahip olduğumuzu bilmek önemlidir.

* Nominal veri: Gözlemler kategoriler halinde gruplanmıştır, değişik karakterlere sahiptir ölçülemez veya gruplanamaz. Cinsiyet, göz rengi, kan grubu gibi. Uygun istatistiksel testler sınırlıdır, ancak frekans dağılımları ve çapraz tablolar uygulanabilir.
* Ordinal veri: Gözlemler kategoriler halinde gruplanmıştır ve azalan veya artan derecelerine göre sıralanabilir. Örneğin hastalığın derecesi hafif, ota ve ağır şeklinde gruplanabilir. Uygun istatistiksel teknikler medyan, eşdeğer ayrımlar arası aralıklar, yüzdeler, dağılım serbestisi olan testlerdir.
* Scale Ölçek veri (Interval ve Ratio veri): Ölçek veride, değerler arasındaki biriler sayısal olarak eşittir. Örneğin uzunluk ve ağırlık gibi. Birçok istatistik teknik bu türden veriler için uygundur. Örneğin: ortalamalar, standart sapmalar, t-testleri vb.

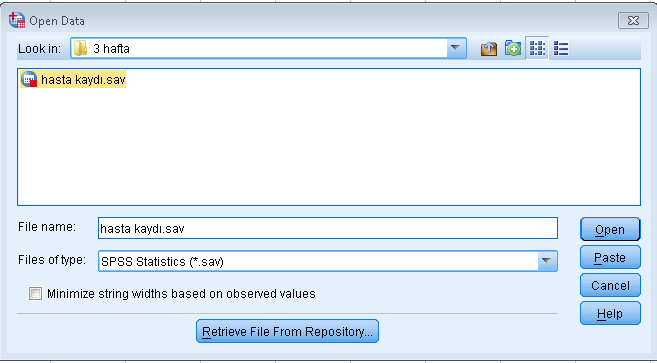
1. **Verileri Özetlemek**

Özet istatistikler lokasyon ölçümü (örneğin: ortalama, mod ve medyan) ve dağılımın ölçülmesi (Aralık, eşdeğer ayrımlar arasındaki aralıklar ve standart sapmadır).

**2.1. Lokasyon ölçülmesi**

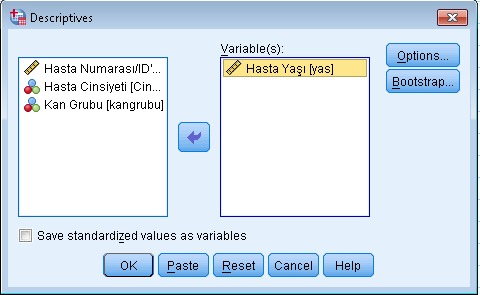
Belirli bir tabloya ortalama değerinin verilmesidir. Ortalama tüm değerlerin toplanması ve değer sayısına bölünmesi ile elde edilir. Örneğin 7, 3, 11, 12, 9, 14 sayılarının ortalaması 56/6 = 9.3’tür.

SPSS’te bir önceki hasta bilgileri örneğimizi hatırlayacak olursak, Ortalama değeri bulmak için; önce SPSS programını çalıştıralım ve veri setini açalım. SPSS programını çalıştırmak için Windows menüsünden tüm programlar ve SPSS.exe executive dosyasını çalıştıralım. Dosyayı açmak için, File>Open>Data komutunu seçerek verimizi kaydettiğimiz yerden çağıralım:



Ok tuşuna basıldığında hasta kaydı.sav SPSS dosyası yüklenecektir. Ortalama değeri bulmak için

* Analyze>Descriptive Statistics>Descriptives menü seçeneklerini takip edelim
* Açılan diyalog kutusunda hasta Yası [Yas] değişkenini seçerek  tuşuna basarak sol tarafa atalım. Diyalog kutusu aşağıdaki gibi görünmelidir.



* Bunun arkasından OK tuşuna basalım ve aşağıda görülen SPSS output’u (çıktısını) alalım. SPSS’de analiz sonuçları otomatik olarak Output1[Document1] penceresine kaydedilir ve bu pencere aşağıda gösterilen menü çubuğu üzerinden görülebilir.



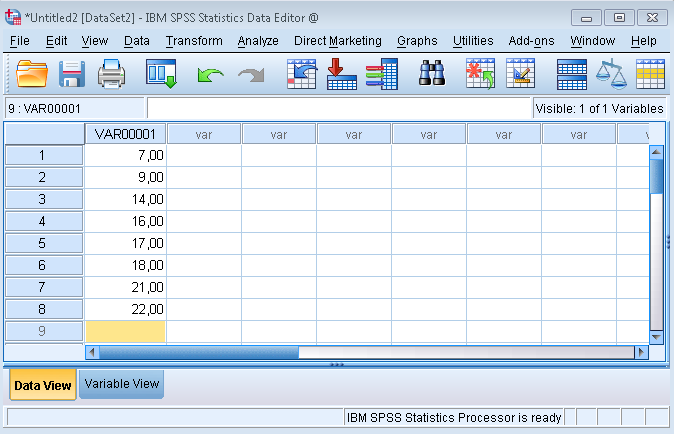
| **Descriptive Statistics** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | N | Minimum | Maximum | Mean | Std. Deviation |
| Hasta Yaşı | 10 | 19,00 | 65,00 | 42,0000 | 16,19328 |
| Valid N (listwise) | 10 |  |  |  |  |

Yukarıdaki Tablo Standart SPSS çıktısıdır ve bir tablo halinde verilmiştir. Hasta yaşı ile ilgili bazı tanımlayıcı istatistikleri vermektedir. N örnek sayısıdır. Örneğimiz 10 hasta içermektedir. Valid N Listwise eksik veri olup, olmadığımızı listemizde gösterir 10 örneğin 10’u da geçerlidir (Valid) ve analizlere kullanılmış, dâhil edilmiştir. Değerlerin minimum’u 19, maksimumu 65’tir. Yani hastaların yaşları 19 ile 65 arasında değişmektedir. Ortalama değer Mean olarak gösterilmiştir ve değeri 42’dir. Std Deviation Standart Sapma değeridir ve verilerin ortalamadan ne kadar saptığını gösteren bir ölçü birimidir. Bu veri setinde standart sapma 16,19’dur yani verilerimiz ortalama +- standart sapma yani 42 +/- 16,19 kadar bir sapma göstermektedir.

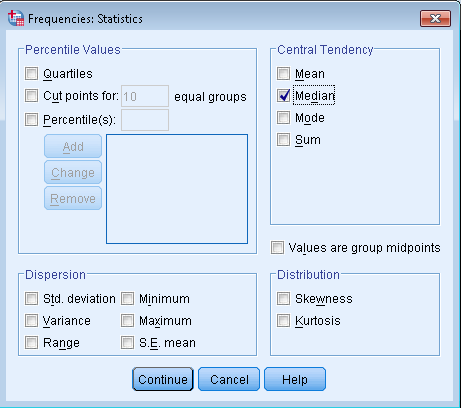
Medyan küçükten büyüğe dizilmiş sayı setinin tam ortasındaki değerdir. Eğer eşit miktarda sayı varsa, tam ortadaki 2 değerin ortalamasıdır. 7, 9, 14, 16, 17, 18, 21, 22 sayılarının medyanı (16 + 17)/2 = 16.5’tir. Medyanı elde ederken sayıların bir yarısının medyandan ufak veya eşit, diğer yarısının da medyandan büyük veya eşit olduğunu biliriz.

SPSS’te medyan değerini bulmak için histogram menüsünden yaralanmamız gerekmektedir. Bunu için yukarıdaki değerleri yeni bir SPSS tablosuna girelim.

* File>New>Data menüsünü çağıralım. Yeni bir SPSS hesaplama tablosu açılacaktır.
* Yukarıda yer alan verileri ilk sütuna yukarıdan aşağıya doğru girelim. Bir veriyi girdikten sonra Tab tuşuna basıldığında aşağıdaki satırda yeni bir veri girişi için hazır hale gelecektir SPSS. SPSS Spreadsheet’imiz aşağıdaki gibi gözükecektir



* Analyze>Descriptive Statistics>Frequencies menüsünü takip ederek, frequency diyalog kutusunu açalım.
* VAR00001 değişkenini önce seçelim sonra  tuşu ile sağ taraftaki variable(s) diyalog kutusu içerisine alalım.
* Sağ tarafta yer alan  kutucuğuna basalım ve statistics opsiyonlarını açalım.
* Sağ tarafta yer alan central tendency ölçülerinden median’ı seçelim. Diyalog kutusu aşağıdaki gibi gözükecektir.

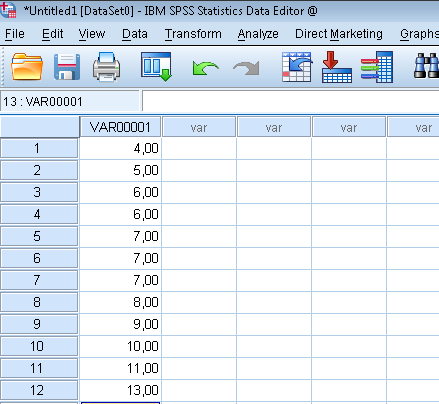


* Continue tuşuna basalım ve ardından OK tuşuna basalım. SPSS çıktılarında median değeri aşağıdaki Tabloda olduğu gibi verilecektir.

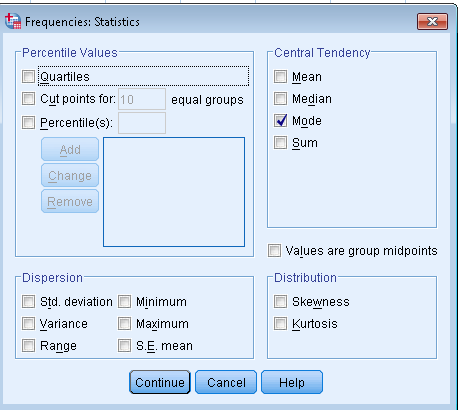
| **Statistics** | | |
| --- | --- | --- |
| VAR00001 | | |
| N | Valid | 8 |
| Missing | 0 |
| Median | | 16,5000 |

Görüldüğü gibi median değeri 16,5’tir ve bizim hesaplarımızla aynıdır.

Mod bir büyükten küçüğe doğru dizilmiş sayı grubunda en çok tekrar eden sayıdır. Örneğin 4, 5, 6, 6, 7, 7, 7, 8, 9, 10, 11, 13 sayı dizisinde en çok tekrar eden sayı 7 olduğu için mod değeri 7’dir. Bu değeri SPSS’te hesaplatmak için önce tüm bu değerleri aşağıda görüldüğü gibi SPSS’e girelim.



* Analyze>Descriptive Statistics>Frequencies menüsünü takip ederek, frequency diyalog kutusunu açalım.
* VAR00001 değişkenini önce seçelim sonra  tuşu ile sağ taraftaki variable(s) diyalog kutusu içerisine alalım.
* Sağ tarafta yer alan  kutucuğuna basalım ve statistics opsiyonlarını açalım.
* Sağ tarafta yer alan central tendency ölçülerinden mod’u seçelim. Diyalog kutusu aşağıdaki gibi gözükecektir.



SPSS çıktısı da aşağıdaki gibi olacaktır.

| **Statistics** | | |
| --- | --- | --- |
| VAR00001 | | |
| N | Valid | 12 |
| Missing | 0 |
| Mode | | 7,00 |

Mod değeri 7 olarak hesaplanmıştır. Aşağıda genel olarak veri setinin frekans dağılımları da verilmiştir.

| **VAR00001** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 4,00 | 1 | 8,3 | 8,3 | 8,3 |
| 5,00 | 1 | 8,3 | 8,3 | 16,7 |
| 6,00 | 2 | 16,7 | 16,7 | 33,3 |
| 7,00 | 3 | 25,0 | 25,0 | 58,3 |
| 8,00 | 1 | 8,3 | 8,3 | 66,7 |
| 9,00 | 1 | 8,3 | 8,3 | 75,0 |
| 10,00 | 1 | 8,3 | 8,3 | 83,3 |
| 11,00 | 1 | 8,3 | 8,3 | 91,7 |
| 13,00 | 1 | 8,3 | 8,3 | 100,0 |
| Total | 12 | 100,0 | 100,0 |  |

Dikkat edilirse en fazla tekrar eden sayı 3 tekrar ile 7’dir ve bu da bize Mod değerini vermektedir. 7 tüm sayıların %25’ine eşdeğerdir.

**2.2. Dağılım Ölçüleri**

Verilerdeki değişkenliği belirleyen bazı ölçütler vardır. Bu ölçütler, aralık, eşdeğerli aralık ve standart sapmadır.

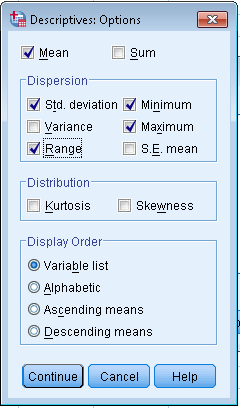
Aralık seri içerisindeki en büyük değerden, en küçük değerin farkıdır. 19, 21, 22, 22, 25, 27, 28, 42 serisinde aralık (range) = 42 – 19 = 23 değeridir. Büyük aralık değeri sayıların en büyüğü ile en küçüğü arasındaki farkın büyük olduğunu gösterir. Ancak aralık yüksek değerlerdeki uç değerlerden fazla etkilenir. Bunu etkiyi önlemek için eşit aralıklı aralık (interquarter range – IQR) kavramı geliştirilmiştir. Bu verilerin %75’lik dilimi ve %25’lik dilimi arasındaki farktır. Yukarıdaki örnekte en uç değerler ayrı tutulursa, aşağı %25, ½\*(21 + 22) = 21.5 ve üst çeyrek %75, ½\*(27 + 28)/2 = 27.5’tur.

SPSS ortamında range’i hesaplamak için öncelikler hasta kaydı. Sav dosyamızı tekrar açalım. Eğer bir önceki alıştırmayı yaparken hasta kaydı.sav dosyasını kapatmadıysak aşağıda program çubuğunda yer alan pencerelerde gözükecektir.



Üzerine tıklayıp aktif hale getirelim. Eğer dosyayı kapatmış isek File>Open>Data menüsünü takip ederek dosyamızı kayıtlı olduğu yerden çağıralım.

* Analyze> Descriptive Statistics>Descriptives menüsünü takip edelim
* Eğer hali hazırda bir önceki örnekteki çalışmamızdan kalarak hasta yası[Yas] sol taraftaki variable(s) kutucuğunda değilse sol tarafa alalım.
* Sağ tarafta yer alan “options” kutucuğunu tıklayalım
* Dispersion seçenekleri arasında yer alan Range kutucuğunu tıklayalım. Kutucuğumuz aşağıdaki gibi görünecektir



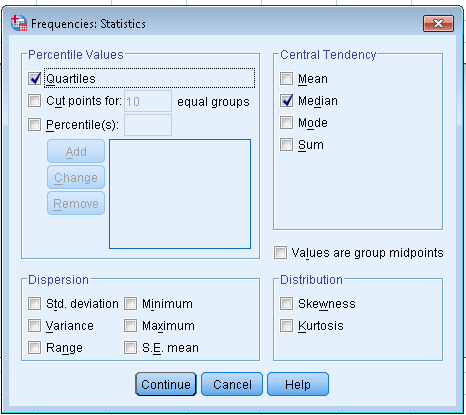
* Continue ve OK kutucuklarını tıklayalım. SPSS’in standart çıktısı aşağıdaki gibi olacaktır.

| **Descriptive Statistics** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | N | Range | Minimum | Maximum | Mean | Std. Deviation |
| Hasta Yaşı | 10 | 46,00 | 19,00 | 65,00 | 42,0000 | 16,19328 |
| Valid N (listwise) | 10 |  |  |  |  |  |

Yukarıdaki Tablo’dan da görüldüğü gibi range değeri 65,00 ile 19,00 arasındaki fark olarak 46,00 olarak hesaplanmıştır.

Interquarter Range değerini hesaplamak için bir önceki örnekte kaydetmediğimiz ve bu yüzden untitled2 olarak menü çubuğumuzda görülen median hesapladığımız veri setini üzerine tıklayarak çağıralım.

* Analyze>Descriptive Statistics>Frequencies seçeneklerini seçelim
* Açılan diyalog kutusunda VAR0001 sağ taraftaki variable(s) kutucuğu içerisinde değilse değişkeni seçtikten sonra ortadaki oka tıklayıp sol tarafa alalım
* Açılan diyalog kutusundan sağ tarafta yer alan “Statistics” kutucuğunu tıklayalım ve percentage values seçenekleri altında yer alan Quartiles seçeneğini tıklayalım. Menü çubuğu aşağıdaki gibi gözükecektir.



* Continue ve Ok kutucukları tıklandığında Quartiles’lar aşağıdaki Tabloda yer aldığı gibi SPSS tarafından üretilecektir.

| **Statistics** | | |
| --- | --- | --- |
| VAR00001 | | |
| N | Valid | 8 |
| Missing | 0 |
| Median | | 16,5000 |
| Percentiles | 25 | 10,2500 |
| 50 | 16,5000 |
| 75 | 20,2500 |

İlk %25’lik dilimin ortalama değeri Tablodan görüldüğü gibi 10,25, %50’lik dilim 16,50 ki bu ortalama değere eşittir ve %75’lik kısım ise 20,25 olarak hesaplanmıştır.

Değişkenliğin bir diğer önemli ifadesi standart sapmadır. Bu değer değişkenlerin ortalamadan sapma derecelerini belirtir. Değişkenlerin ortalamadan farklarının, ortalamasının karekökü olarak hesaplanır. Standart sapmayı hesaplamak için öncelikle alt tarafta yer alan menü çubuğundan hasta kaydı.sav dosyasını seçip, aktif hale getirmemiz gerekmektedir.

* Analyze>Descriptive Statistics>Descriptives menüsünü seçerek descriptives yani tanımlayıcı istatistikler diyalog kutusunu açalım
* Eğer hasta yası[Yas] değişkeni hali hazırda sağ tarafta değilse, önce seçelim sonra ok tuşuna basarak sağ tarafa variable(s) değişkenleri arasına alalım.
* Ok tuşuna basalım. Daha önce karşımıza çıkan Tanımlayıcı istatistik çıktıları SPSS tarafında üretilecektir ve aşağıda verilmiştir.

| **Descriptive Statistics** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | N | Range | Minimum | Maximum | Mean | Std. Deviation |
| Hasta Yaşı | 10 | 46,00 | 19,00 | 65,00 | 42,0000 | 16,19328 |
| Valid N (listwise) | 10 |  |  |  |  |  |

Tablonun en son sütununda yer alana Standard Deviation, Standart Sapmadır ve değeri görüldüğü gibi 16,19’tur. Bunun anlamı hasta verilerimizde yer alan 10 hastanın yaşları ortalama çevresinde +/- 16,19 yıllık bir sapma göstermektedir.

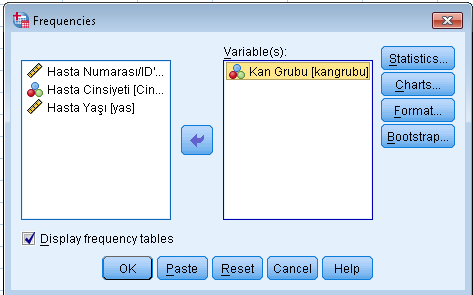
**2.3. Veriyi Sunmak**

Veriler bar grafikleri, histogramlar ve dağılım tabloları ile sunulurlar.

Dağılım Tabloları

Dağılım Tabloları verileri sunmanın basit ve etkin bir yoludur. Özellikler veriler değişik türden kategoriler halinde inceleniyorsa (nominal veya ordinal veriler) bu tür gösterim oldukça uygun olabilmektedir. Bir önceki örneğimizden hatırlayacağımız hasta no ve yaş gibi sürekli değişkenler için dağılım tabloları uygun olmamaktadır. Ancak Kan grubu veya cinsiyet için oldukça uygundur. SPSS’de frekans tablosu oluşturmak için öncelikler

* Analysis>Desciptive Statistics>Frequencies menüsü takip edilir
* Açılan diyalog kutusundan Kan grubu [Kangrubu] değişkeni sağ tarafa alınır ve kutu aşağıdaki gibi gözükür



* Ok tuşuna basılır ve SPSS çıktısı alınır. Outpute penceresinde frekans dağılımı tablosu aşağıda verildiği gibi oluşturulacaktır.

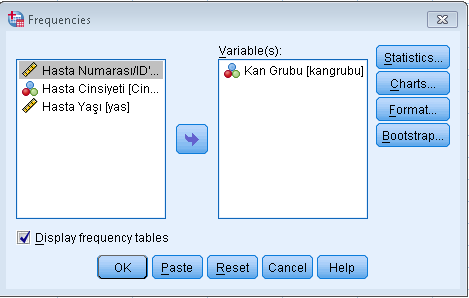
| **Kan Grubu** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 0 | 4 | 40,0 | 40,0 | 40,0 |
| A | 3 | 30,0 | 30,0 | 70,0 |
| AB | 1 | 10,0 | 10,0 | 80,0 |
| B | 2 | 20,0 | 20,0 | 100,0 |
| Total | 10 | 100,0 | 100,0 |  |

Hatırlanacağı üzere bir önceki derste değişkenlerin yer alacağı değerler (value’lar) girildiği için kan grupları sol ilk sütunda 0, A, AB ve B grubu olarak görülmektedir. Tabloyu yorumlayacak olursak; 0 grubu kana sahip olan 4 kişi veri setimizde mevcuttur ve bu da tüm katılımcıların veya örneğimizde hastaların %40’ına tekabül etmektedir. A grubu kana sahip 3 kişi %30, AB grubu kana sahip 1 kişi %10 ve B grubu kana sahip 2 kişi %20’lik toplam hasta oranlarını vermektedir. Burada dikkat edilecek bir nokta “percent ve “valid percent” sütunlarının birbirine eşit olmasıdır. Çünkü veri setimizde hiçbir kayıp (missing) değer yoktur. Eğer veri setimizde kayıp (missing) değerler olsa idi, kayıp değerler düşüldükten sonra yeniden yüzdeler hesaplanacak ve bu da “Valid Percent” sütununda gösterilecekti. Bu durumda bizim dikkate almamız gereken değer “Valid Percent” sütunu olacaktı. En sonda ise “Cumulative Percent” sütununda kümülatif toplamları görmekteyiz ve bunların nihai toplamı da %100’e denk gelmektedir.

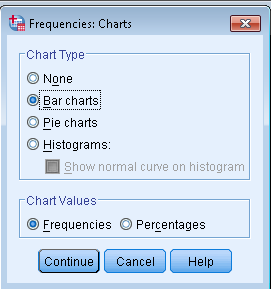
Bar (Çubuk) Grafiği

Bar Grafiği frekans dağılımındaki her bir değişken veya dağılımın sayısını gösteren görsel bir grafiktir. Bar grafiğini elde etmek için frekans tablosu menüsünü açmamız gerekmektedir.

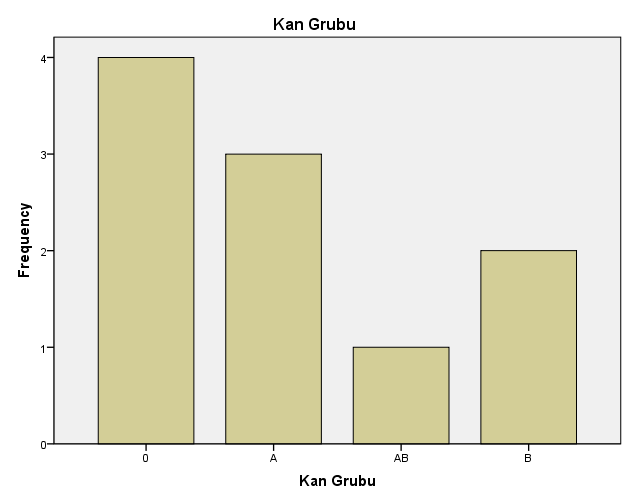
* Analyze>Descriptive Statistics>Frequencies menüsü takip edilir
* Açılan menüden Kan Grubu[Kangrubu] değişkeninin sağ tarafta olduğuna emin olunur ve sağ tarafta bulunan “Charts” kutucuğu aşağıda gösterildiği gibi tıklanır.



* Açılan Charts diyalog kutusundan “Chart Type” Seçeneklerinden “Bar Charts” aşağıda gösterildiği gibi seçilir



* Continue ve Ok tuşlarına basıldığında SPSS bize aşağıda verildiği gibi bar Chart’ı oluşturacaktır.

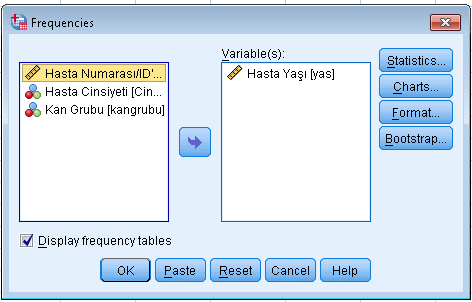


Burada yine en uzun çubuk bize 0 grubundan en fazla sayıda tam olarak 4 hasta olduğunu göstermektedir. Sonra A grubundaki 3 hasta, B grubundaki 2 hasta ve AB grubundaki 1 hasta ile toplam 10 kişinin kan grubu dağılımı verilmiştir.

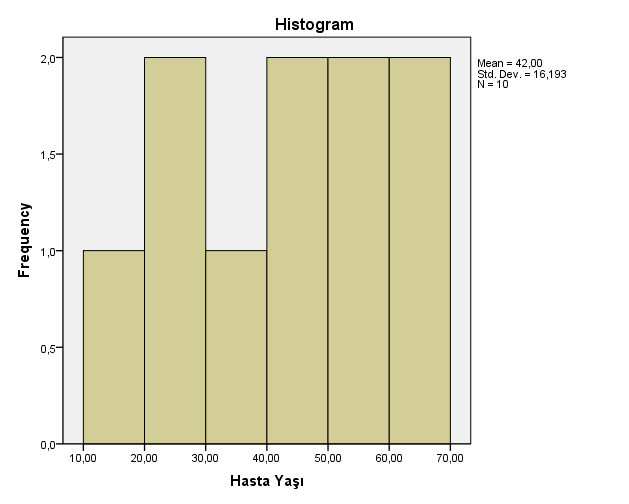
Histogram

Histogram Bar Grafiğine benzese de, her bir çubuk değerlerin aralığını belirtmektedir. Histogramı kullanmak için sürekli değişkenlere ihtiyaç vardır. Örneğimizde hastaların yaşının histogram grafiğini almak istersek,

* Analyze>Descriptive Statistics>Frequencies menüsü takip edilir
* Sağ taraftaki variable(s) kutucuğunda eğer Kan Grubu[Kangrubu] değişkeni varsa, seçilip sol tarafa alınır ve hasta Yaşı [Yas] değişkeni aşağıda görüldüğü gibi sağ tarafa alınır

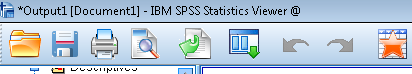


* Sağ tarafta yer alan Charts kutucuğu tıklanır ve açılan diyalog kutucuğundan histogram seçilir
* Aşdağıdaki gibi bir çıktı output ekrarnında görülür

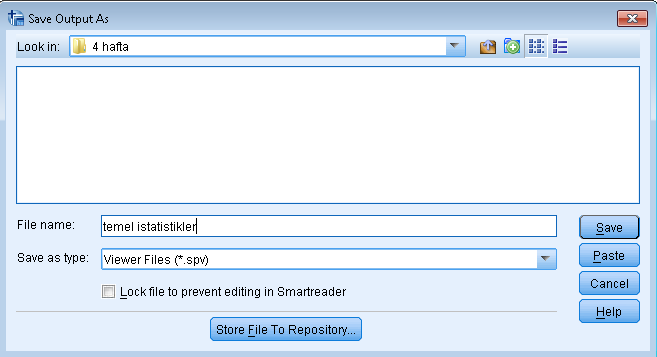


Veri setinde yaşları 10 ile 20 arasında 1 kişi, 20 ile 30 arasında 2 kişi, 30 ile 40 arasında 1 kişi, 40 ile 50 arasında, 50 ile 60 ve 60 ile 70 arasında 1’er kişinin olduğu histogramımızdan görülmektedir.

SPSS’den çıkmadan önce output dosyasını bize kaydedip kaydetmeyeceğimiz sorulacaktır. Aşağıda gösterildiği gibi output penceresi açıkken, kayıt düğmesine tıklayarak buraya kadar yaptığımız tüm işlemlerin SPSS çıktılarını kaydedebiliriz.



Aşağıda gösterildiği gibi çıktımıza temel istatistikler adını verip .spv uzantısı ile SPSS çıktılarını bilgisayarımızda istediğimiz yere kaydedebiliriz. Bu çıktıları çağırmak için File>Open>Output menüsünden kaydettiğimiz yerden .spv dosyasını bulup, geri çağırmamız yeterli olacaktır.



**Değerlendirme Soruları**

1. SPSS’te kaç türlü ölçek türü vardır? Açıklayınız.
2. SPSS’te Mod, medya ve ortalama nasıl hesaplanır?
3. SPSS’te histogram ve Bar Charts’ı nasıl elde ederiz\_.

**Kaynakça**

Spiers, N., Manktelow, B. Ve Hewitt, M. J.(2009), Practical Statistics Using SPSS, National Institude for Health Research NHS, England.

Field, A. (2005), Discovering Statistics using SPSS, SAGE, London.