# Verilerin grafiklerle gösterilmesi

Amaç: Bu konu sonunda okuyucu verilerin grafiklerle gösterilmesi konusunda bilgi sahibi olmalı ve SPSS kullanarak verileri grafiklerle gösterebilmelidir.

Hedefler: Bu konu sonunda okuyucuların aşağıdaki hedeflere ulaşması beklenmektedir:

* Grafiklerde bulunması gereken temel özellikleri sayabilmek
* Temel grafik çeşitlerini sayabilmek
  + Bar grafik
  + Pasta dilimi
  + Histogram
  + Dot plot
  + Boxplot (saplı kutu grafiği, asansör grafiği)
  + Scatter diagram (saçılma grafiği, nokta grafikler)
* Verilere uygun grafik seçebilmek
  + Tek değişken olduğunda
  + Çok değişken olduğunda
  + Numerik veri olduğunda
  + Kategorik veri olduğunda
* Temel grafikleri SPSS kullanarak yapabilmek
* Frekans dağılımının yönünü belirleyebilmeli
  + Sağa eğimli
  + Sola eğimli

Verilerimizi bilgisayara girdikten sonra öncelikli olarak yapmak isteyeceğimiz şey onları özetlemek ve bir şekilde “hissedilebilir” hale getirmektir. Bunun en güzel yolu da tablolar ve grafikler hazırlayarak özet istatistikler oluşturmaktır. Grafiklere bakarak daha hipotez testlerini uygulamadan ve önemlilik durumuna bakmadan durum hakkında bir fikir edinebiliriz.

## Grafik özellikleri

Grafik hazırlarken aşağıda belirtilen özellikler olmalıdır.

* Ne, nerede, ne zaman sorularına yanıt veren, önemli ilişkileri belirten bir başlığı, sıra numarası olmalıdır.
* Dikey ve yatay eksenlerin gösterdikleri değişkenler ve varsa bunların birimleri açık biçimde belirtilmelidir.
* Genellikle bağımlı değişken dikey, bağımsız değişken yatay eksende gösterilir.
* Grafik çok sayıda çizgi veya eğrilerle karmaşık hale dönüştürülmemelidir.
* Grafikler genellikle önemli verilerin özetlerini, vurgulayan noktalarını göstermek amacıyla düzenlenirler. Bu bakımdan çok ayrıntı vermek hatalıdır.

## Frekans dağılımı

**Ampirik (deneysel; empirical) frekans dağılımı**, her bir değişkenle ilgili gözlemlerin, bunların sınıflarının, veya kategorilerinin gözlem **frekansı** (sıklığı) açısından gösterilmesidir. Frekans yerine **relatif (nisbi) frekans** kullanmamız halinde iki ya da daha fazla grup arasında frekans dağılımlarını karşılaştırabiliriz.

## Frekans dağılımlarının gösterilmesi

Frekans dağılımları kategorik veriler veya bazı tam sayılı numerik veriler için kullanılır; verilerin görsel olarak sunulmasını sağlar.

* **Bar veya sütun grafikleri**: her bir kategori için ayrı bir yatay veya dikey sütun çizilir. Sütunun boyu değişkenin ilgili kategorideki sıklığı (frekansı) ile orantılıdır. Sütunlar arasındaki boşluklar değişkenlerin kategorik veya tam sayılı olduğunu gösterir. Örn.:

[*www.aile.net/agep/istat/08\_09/diyabet.sav*](http://www.aile.net/agep/istat/08_09/diyabet.sav) *dosyasını SPSS ile açınız. “marital status değişkeninin frekans dağılımına bakalım. sütun grafiğini çıkaralım. Analyze>Descriptive Statistics>Frequencies>[“Marital status” değişkenini “Variable(s)” alanına geçirelim]>ok*

*Şu çıktıları elde ederiz:*

**Statistics**

Marital status

|  |  |  |
| --- | --- | --- |
| N | Valid | 420 |
| Missing | 10 |

*Bu değişken için 420 veri girilmiş. 10 adet te boş veri bulunuyor.*

**Marital status**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Single | 24 | 5,6 | 5,7 | 5,7 |
| Married | 335 | 77,9 | 79,8 | 85,5 |
| Divorced | 20 | 4,7 | 4,8 | 90,2 |
| Widow | 41 | 9,5 | 9,8 | 100,0 |
| Total | 420 | 97,7 | 100,0 |  |
| Missing | No response | 3 | ,7 |  |  |
| System | 7 | 1,6 |  |  |
| Total | 10 | 2,3 |  |  |
| Total | | 430 | 100,0 |  |  |

*24 kişi bekarmış (Single). Bu 24 kişi toplamın %5,6’sını, eksik veriler çıkarıldığında geri kalanların ise %5,7’sini oluşturuyor. 3 kişi anketin bu sorusuna cevap vermemiş (no response). SPSS’te “No response” kategorisinin “Missing” başlığı altında 9 olarak kodlandığına dikkat ediniz. Ayrıca 7 cevapsız daha var. Bunlar “System missing” olarak belirtilmiş. Bu verilerin anketin uygulanması aşamasında veya bilgisayara girilmesi aşamasında unutulduğunu varsayabiliriz.*

*Şimdi aynı frekans dağılımını sütun grafiği ile gösterelim:*

*Graphs>Interactive>Bar [X eksenine “Marital status” değişkenini koyalım]>OK*

*Aşağıdaki grafiği elde ederiz:*

**

*Grafiğimizde sütunların üzerinde sayı ve/veya yüzdelerin görülmesini istersek OK butonunu tıklamadan önce sekmelerden gerekli tercihleri yapabiliriz veya daha sonra grafik çıktısının üzerine çift tıklayarak da ayarları değiştirebiliriz.*

* **Pasta grafiği**: daire şeklindeki bir pasta her bir dilimi değişkenin ilgili kategorisinin frekansını temsil edecek şekilde dilimlere ayrılır. Örn.:

[*www.aile.net/agep/istat/08\_09/diyabet.sav*](http://www.aile.net/agep/istat/08_09/diyabet.sav) *dosyasını SPSS ile açınız. “marital status değişkeninin pasta grafiğini yapalım.*

*Graphs>Interactive>Pie>Simple [Slice by kutusuna “Marital status” değişkenini sürükleyelim. Üstteki “Pies” sekmesini tıklayıp “Percent” kutucuğunu işaretleyelim]>OK*

*Aşağıdaki grafiği elde ederiz:*



* **Histogram**: bu grafik de sütun grafiğine banzer. Ancak, sütunlar arasında boşluk yoktur. Sütunlar bir değeri değil, belli aralıktaki değerleri temsil eder. Bu grafik daha çok numerik değişkenlerde verilerin dağılım şeklini incelemek için kullanılır. Örn.:

[*www.aile.net/agep/istat/08\_09/diyabet.sav*](http://www.aile.net/agep/istat/08_09/diyabet.sav) *dosyasını SPSS ile açınız. “Weight” değişkeninin histogram grafiğini çizelim.*

*Graphs>Interactive>Histogram [X eksenine “Weight” değişkenini sürükleyelim. Üstteki “Histogram” sekmesini tıklayıp “Normal curve” kutucuğunu işaretleyelim]>OK*

*Aşağıdaki grafiği elde ederiz:*



*Verilerimizin 75 kg civarında yoğunlaştığını, dağılımın çan eğrisine benzediğini, çan eğrisinin sağ taraftaki kuyruğunun biraz daha uzun olduğunu görüyoruz.*

* **Dot plot**: sık kullanmadığımız bir grafik şeklidir. Her bir gözlem yatay veya dikey eksende bir nokta olarak gösterilir. Büyük veri setlerinde bu grafikleri yorumlamak zor olabilir (noktalar üst üste binecektir).
* **Boxplot (saplı kutu grafiği)**: Asansör grafiği, kutu ve sakal grafiği (box and whisker plot) de denir. “Hata ayıklama ve uç değerler” konusuna bakınız.
* **Scatter diagram** (saçılma grafiği, nokta grafikler): iki ayrı numerik veya ordinal değişkeni grafikle göstermenin bir yolu nokta grafikleridir. Örn.:

[*www.aile.net/agep/istat/08\_09/diyabet.sav*](http://www.aile.net/agep/istat/08_09/diyabet.sav) *dosyasını SPSS ile açınız. “Height” ve “Weight” değişkenlerinin histogram grafiğini çizelim.*

*Graphs>Interactive>Scatterplot [X eksenine “Height” değişkenini, Y eksenine de “Weight” değişkenini sürükleyelim]>OK*

*Aşağıdaki grafiği elde ederiz:*



*Grafikten boy arttıkça ağırlığın da arttığını görebiliyoruz.*

## Frekans dağılımının şekli

Hangi istatistiksel testi uygulayacağımız büyük oranda verilerimizin dağılımına bağlıdır. Genelde verilerimizin dağılımı **unimodaldir**, yani tek bir pik noktası vardır. Bazen **bimodal** (iki ayrı pik noktası) ve **uniform** (tüm veriler eşit) dağılımlar da söz konusu olabilir. Unimodal verilerde verinin en büyük ve en küçük değerler arasında nasıl dağıldığını görmek önemlidir. Verilerimizin özellikle bakmak istediğimiz özellikleri şunlardır:

* **Simetrik dağılım**: bir orta değer etrafında eşit olarak dağılır. Histogram grafiğinde çan eğrisinin iki tarafı da simetriktir.
* **Sağa eğimli** (pozitif eğimli): histogram grafiğinde çan eğrisinin kuyruğu sağa doğrudur.
* **Sola eğimli** (negatif eğimli): histogram grafiğinde çan eğrisinin kuyruğu sola doğrudur.

# Merkezi dağılım (yığılma) ölçütleri – the average

Amaç: Bu konu sonunda okuyucu merkezi dağılım ölçütleri hakkında bilgi sahibi olmaları ve SPSS kullanark merkezi dağılım ölçütlerini hesaplayabilmeleri amaçlanmıştır.

Hedefler: Bu konu sonunda okuyucuların aşağıdaki hedeflere ulaşması beklenmektedir:

* Merkezi eğilim ölçütlerini sayabilmeli
* SPSS kullanarak merkezi eğilim ölçütlerini hesaplayabilmeli
* SPSS kullanarak histogram grafiği çıkarabilmeli ve dağılımın durumunu değerlendirebilmeli
* Sağa deviye dağılıma sahip değişkenlere logaritmik dönüşüm uygulayabilmeli
* Ağırlıklı ortalamayı açıklayabilmeli
* SPSS kullanarak ağırlandırma yapabilmeli

Elde ettiğimiz verileri bir şekilde özetlemeden verinin geneli hakkında fikir sahibi olamayız. Grafiklerle gösterme (ders 5) iyi bir özetleme yöntemi olabilir. Verinin belli özelliklerini hesaplayarak da genel bir fikir sahibi olabiliriz. Verilerimizi temsil edece bir değer bulabilir ve verilerin bu değer etrafındaki dağılımını hesaplayabiliriz. “merkezi dağılım ölçütleri” ve “yaygınlık ölçütleri” olarak ifade edilen bu ölçütler bu bölümde ve sonraki bölümde işlenecektir.

## Aritmetik ortalama

Aritmetik ortalamaya sadece “ortalama” da denir; bütün verilerin toplanması ve veri sayısına bölünmesiyle elde edilir.



Formülde aritmetik ortalama x üzerinde çizgi ile gösterilir. Bu formül Yunancadaki Sigma toplama işareti kullanılarak

 veya şeklinde gösterilebilir. Örn.:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Birey no* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |
| *Yaş* | *10* | *13* | *15* | *16* | *16* | *18* | *19* | *20* | *20* |

*Yukarıdaki veri setinde yaşın aritmetik ortalaması (10+13+15+16+16+18+19+20+20)/9 = 16,33’tür.*

## Ortanca

Verilerimizi büyükten küçüğe doğru sıraladığımızda ortadaki değere **ortanca** denir. Eğer veri adedimiz çift sayı ise ortadaki iki değerin ortalaması alınır. Örn.:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Birey no* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |
| *Yaş* | *10* | *13* | *15* | *16* | *16* | *18* | *19* | *20* | *20* |

*Yukarıdaki veri setinde yaşın ortanca değeri 5 numaralı bireyin yaşı olan 16’dır.*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Birey no* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* |
| *Yaş* | *10* | *13* | *15* | *16* | *16* | *18* | *19* | *20* | *20* | *21* |

*Yukarıdaki veri setinde ise 10 adet veri olduğundan ortanca 5 ve 6. Verilerin toplanıp ikiye bölünmesiyle bulunur: (16+18)/2=17*

## Mod

Mod, sık kullanılmayan bir merkezi dağılım ölçütüdür. Veri setinde en fazla tekrar eden değişkene mod denir. Veri setinde birden fazla mod olabilir. Her verinin sadece bir kez tekrarlaması halinde ise mod yoktur. Örn.:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Birey no* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* |
| *Yaş* | *10* | *13* | *16* | *16* | *16* | *18* | *19* | *20* | *20* | *21* |

*Yukarıdaki veri setinde mod 16’dır.*

## Geometrik ortalama

Verilerimizin eğimli olması halinde artimetik ortalamayı kullanmak uygun değildir. Verilerin sağa eğimli olması haline (histogram grafiğinde çan eğrisinin kuyruğu sağa doğru) verilerin tek tek logaritmasını (10 tabanına veya *e* tabanına göre) alırsak elde edeceğimiz yeni veri seti simetrik hale gelebilir. Bu logaritma değerlerinin artimetik ortalamasını alabiliriz. Tabiiki verinin orijinal birimine geri dönmesi için tekrar dönüşüm yapmak (antilog) gerekir. Bu yeni değere ise **geometrik ortalama** denir. Genelde geometrik ortalama ortancaya yakın ve artimetik ortalamadan daha küçük bir değer olur. Örn.:

[*www.aile.net/agep/istat/08\_09/diyabet.sav*](http://www.aile.net/agep/istat/08_09/diyabet.sav) *veri setinde “Weight” değişkenini incelediğimizde histogram grafiğinde çan eğrisinin kuyruğunun sağa doğru olduğunu (sağa eğimli) görmüştük. Şimdi “weight” değişkeninin logaritmasını alarak yeni bir değişken oluşturalım:*

*Transform>Compute variable>[“Target Variable” alanına “LogWeight”, “Numeric Expression” alanına ise “LG10(weight)” yazalım]>OK*

*SPSS veri setimizde “LogWeight” adında yeni bir değişken oluşacaktır. Şimdi bu değişkenin histogram grafiğine bakalım:*

*Graphs>Interactive>Histogram [X eksenine “LogWeight” değişkenini sürükleyelim. Üstteki “Histogram” sekmesini tıklayıp “Normal curve” kutucuğunu işaretleyelim]>OK*

*Aşağıdaki grafiği elde ederiz:*

**

*Görüldüğü gibi çan eğrisi simetrik hale gelmiştir.*

*“Weight” değişkenimizin artimetik ortalamasını hesaplayalım:*

*Analyze>Descriptive Statistics>Descriptives [“Weight” değişkenini “Variable(s)” alanına geçirelim]>OK*

*Aşağıdaki çıktıyı elde ederiz:*

**Descriptive Statistics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | N | Minimum | Maximum | Mean | Std. Deviation |
| Weight | 424 | 33,0 | 160,0 | 74,266 | 15,1381 |
| Valid N (listwise) | 424 |  |  |  |  |

*Görüldüğü gibi “Weight” değişkeninin artimetik ortalaması 74,266 kg’dır.*

*Şimdi de “LogWeihgt” değişkeninin artimetik ortalamasını alalım:*

*Analyze>Descriptive Statistics>Descriptives [“LogWeight” değişkenini “Variable(s)” alanına geçirelim]>OK*

*Aşağıdaki çıktıyı elde ederiz:*

***Descriptive Statistics***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *N* | *Minimum* | *Maximum* | *Mean* | *Std. Deviation* |
| *LogWeight* | *424* | *1,52* | *2,20* | *1,8625* | *,08428* |
| *Valid N (listwise)* | *424* |  |  |  |  |

*Görüldüğü gibi “LogWeight” değişkeninin artimetik ortalaması 1,862’dir.*

*Şimdi elde ettiğimiz değeri klinik açıdan yorumlayabilmek için Weight değişkeninin birimine geri çevirmemiz gerekiyor. 1,862 değerinin anti logaritmasını almalıyız. Antilog (1,862) = 101,862 = 72,777 kg.*

*“Weight” değişkeninin ortanca ve mod değerlerini SPSS ile hesaplamak için:*

*Analyze>Descriptive Statistics>Frequencies [“Weight” değişkenini “Variable(s)” alanına geçirelim. “Statistics” butonuna tıklayalım. “Central tendency” başlığının altından “Median” ve “Mode” kutucuklarını işaretleyelim]>Continue>OK*

*Not: artimetik ortalamayı yukarıdaki menüde “Mean” kutucuğuna işaret koyarak ta hesaplayabiliriz.*

*Aşağıdaki çıktıyı elde ederiz:*

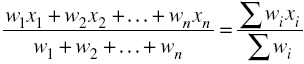
***Statistics***

*Weight*

|  |  |  |
| --- | --- | --- |
| *N* | *Valid* | *424* |
| *Missing* | *6* |
| *Median* | | *72,750* |
| *Mode* | | *70,0* |

## Ağırlıklı ortalama

Bir değişkenin bazı değerlerinin diğerlerinden daha önemli olması halinde **ağılıklı ortalama** kullanırız. Örneklemimizdeki her bir değere bir katsayı veririz. Her bir değerle katsayısını çarpıp bunları toparız. Sonra değerlerin toplamına böleriz.



*Örn. İlimizdeki hastanelerin günlük taburcu sayılarını inceliyoruz. Değişkenimiz “Bu gün hastanenizden kaç hasta taburcu oldu?” şeklinde olsun. İlimizdeki 3 hastane için elde ettiğimiz veriler aşağıdadır:*

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Hatane 1* | *Hastane 2* | *Hastane 3* |
| *Taburcu edilen hasta* | *20* | *5* | *50* |

*3 nolu hastanenin en fazla hasta taburcu ettiğini görüyoruz. Ortalama taburcu sayısı 25’dir. Bu hastanelerin yatak kapasitelerini bilmeden iş yoğunlukları hakkında fikir sahibi olamayız. Hasta kapasiteleri şu şekilde olsun:*

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Hatane 1* | *Hastane 2* | *Hastane 3* |
| *Taburcu edilen hasta* | *20* | *5* | *50* |
| *Yatak kapasitesi* | *50* | *50* | *400* |

*Taburcu sayılarını yatak kapasiyesine göre ağırlıklandırmamız halinde daha geçerli bir fikir elde edebiliriz. Formülü uygulayacak olursak:*

*(20x50 + 5x50 + 50x400)/(50+50+400) =* ***42,5 taburcu****.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Hatane 1* | *Hastane 2* | *Hastane 3* | *Ortalama* |
| *Taburcu edilen hasta* | *20* | *5* | *50* | *25* |
| *Yatak kapasitesi* | *50* | *50* | *400* | *166,6* |
| *Ağırlıklı taburcu* | *66,6\** | *16,6* | *20,8* |  |
| *\* 20 x 166,6 / 50* | | | | |

*Sonuç olarak ağırlıklı ortalamanın (44 kişi) ilk başta görünenden (25 kişi) çok daha fazla olduğunu ve 3 nolu hastanenin en az yoğunlukta çalıştığını görüyoruz.*

*Şimdi de ağırlıklandırmayı SPSS ile yapalım:*

*Çocuk sayısının yaş faktöründen etkilendiğini ve yaş ilerledikçe daha fazla çocuk sahibi olunacağını biliyoruz.* [*www.aile.net/agep/istat/08\_09/diyabet.sav*](http://www.aile.net/agep/istat/08_09/diyabet.sav) *veri setinde “children” (çocuk sayısı) değişkenini yaşa göre ağırlıklandıralım. Ağırlıklandırma yapmadan önce “children” değişkeninin aritmetik ortalamasının 6,38 olduğuna dikkat ediniz.*

*Data>Wieght Cases>Weight cases by>[“Age” değişkenini “Frequency Variable” alanına geçiriniz]>OK.*

*Şimdi çocuk sayısının (“children”) aritmetik ortalamasını aldığımızda 6,61 olduğunu görüyoruz.*

*Bu işleme “Yaşa göre düzeltilmiş çocuk sayısı” da denir. Uluslararası istatistiklerde ölüm oranları gibi istatistikler nüfusa veya başka değişkenlere göre düzeltilerek (ağırlıklandırılarak) verilir.*

Ortalamaların olumlu ve olumsuz yönleri aşağıdaki tabloda görülmektedir:

|  |  |  |
| --- | --- | --- |
| **Ortalama türü** | **Olumlu** | **Olumsuz** |
| Aritmetik ortalama | * tüm değerler kullanılır * cebirsel olarak tanımlanmış ve matematiksel olarak kullanılabilir * bilinen örneklem dağılımı (bakınız Bölüm: veri dönüştürme) | * uç değerlerden etkilenir * eğimli verilerden etkilenir |
| Medyan (Ortanca) | * uç değerlerden etkilenmez * eğimli (skewed) veriden etkilenmez | * bilginin büyük bir kısmı ihmal edilir * cebirsel olarak tanımlanmamıştır * örneklem dağılımından etkilenir |
| Mod | * kategorik veriler için kolayca saptanabilir | * bilginin büyük bir kısmı ihmal edilir * cebirsel olarak tanımlanmamıştır * örneklem dağılımı bilinmez |
| Geometrik ortalama | * geri dönüştürme yapılmadan önce aritmetik ortalamayla aynı avantajlara sahiptir * sağa doğru eğimli veriler için uygundur | * sadece log dönüşüm simetrik bir dağılım oluşturuyorsa işe yarar |
| Ağırlıklı ortalama | * artimetik ortalamayla aynı avantajlara sahiptir * her bir gözleme nisbi bir önem verilir * cebirsel olarak tanımlanmıştır | * ağırlık bilinmeli veya hesap edilmelidir |

Ders notunun yer aldığı web sitesi:

* [http://aile.atauni.edu.tr](http://aile.atauni.edu.tr/)