# Veri çeşitleri

Hedefler: Bu konu sonunda okuyucuların aşağıdaki hedeflere ulaşması beklenmektedir:

1. Aşağıdaki terimleri tanımlayabilmeli
2. Veri
3. Değişken
4. Parametre
5. Toplum
6. Örneklem
7. İstatistik
8. Censored (durdurulmuş) veri
9. Verileri sınıflandırabilmeli
10. Aşağıdaki türetilmiş veri çeşitlerini açıklayabilmeli
11. Yüzde
12. Oran
13. Orantı
14. Hız

*2500 öğrencisi olan bir okuldaki [evren ya da* ***toplum****] öğrencilerden rastgele seçilen 250 öğrencide [****örneklem****] boy ve ağırlık* ***değişken****leri hakkında* ***veri*** *toplamaya karar verdiğimizi düşünelim. Öğrencilerin boy ve kilolarını inceleyerek şişmanlık durumlarının* ***istatistik*** *analizini yapmayı hedefleyebiliriz.*

Araştırma yapmadaki öncelikli amaç bir konuyla ilgili veri (data) toplamaktır. Türkçe sözlükte “veri”, “Bir araştırmanın, bir tartışmanın, bir muhakemenin temeli olan ana öge, done” olarak tanımlanmaktadır.

Verilerimizi elde ettiğimiz kaynağın tamamına “**toplum**” veya “**evren**” denir. Toplumun hakkında veri topladığımız özelliklerine ise “**değişken**” adı verilir. Değişkenleri birbirine olan etkileri açısından **bağımlı değişken** ve **bağımsız değişken** olarak sınıflandırabiliriz.

İnceleyeceğimiz değişkeni toplumun tamamında araştırmak için kaynak, zaman ve imkanlarımız çoğunlukla yetersiz olacağından genelde toplumu temsil edebilecek bir grup birey yani “**örneklem**” hakkında veri toplarız. Eğer incelediğimiz değişken hakkında toplumun tamamından veri toplamışsak o zaman “değişken” ifadesi yerine “**parametre**” kullanmak daha uygun olur. Değişken yerine parametre terimini kullanmak sık yapılan yanlışlardandır.

Elde ettiğimiz verileri yöntemli bir biçimde toplayıp sayı olarak belirtme, özetleme, analiz etme ve sonuç çıkarma işine ise “**istatistik**” denir.

Değişkenler **kategorik** ve **numerik** olarak ikiye ayrılır.

Kategorik değişkenlerde her birey belli sınıflardan birine dahil edilebilir. Bu değişken türü de kendi içerisinde **Nominal** (Atanmış) ve **Ordinal** (Sıralı) olmak üzere ikiye ayrılır. Nominal değişkenlerde hangi kategorinin önce geleceği araştırmacının atamasına bağlıdır (1=kadın, 2=erkek olabileceği gibi 1=erkek, 2=kadın da olabilir). Sıralı değişkenlerde ise küçükten büyüğe, ya da büyükten küçüğe doğru bir sıra vardır (1=zayıf, 2=normal, 3=kilolu, 4=şişman, 5=aşırı şişman gibi). Belli bir sıralama (büyükten küçüğe gibi) vardır.

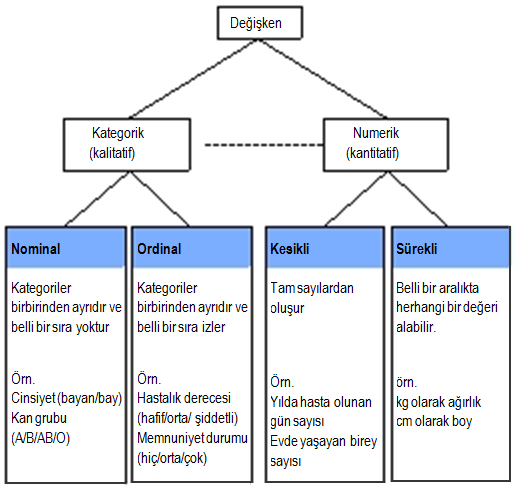
Bir kategorik değişkenin sadece iki özelliği olması halinde ona binary (**ikili**) veya dichotomous (**dikotom**) değişken denir (evet/hayır; canlı/ölü gibi).

Değişkenimiz, sınıflardan oluşmak yerine belli sayıları alan numerik bir değişken de olabilir. Bu durumda da alınan değer **tam sayılı** (kesikli-discret) (Kaç kez doktora gittiniz? [4 kez] Hastane kaç yataklı? [60 yataklı] gibi) veya **sürekli** (continuous) (Boy [165,5 cm], Hemoglobin değeri [14,3 mg/dl] gibi) olabilir.

Değişken çeşitlerini ayırdedebilmek kullanılacak istatistiksel yöntemi seçerken ve SPSS’te analiz yaparken önemli olacak. Kategorik değişkenler sayı ve yüzde olarak gösterilirken numerik değişkenler ortalama ve standart sapma gibi ölçütlerle gösterilir.

Kategorik ve numerik verilerin ayırımı genellikle kesin olmakla birlikte bazen sıralı değişkenlerde kategori sayısı fazla olabilir. Bu durumlarda sıralı değişkenler kategorikmiş gibi muamele görebilir. Örneğin ağrı şiddetini araştırdığımız bir ankette 1’den 7’ye kadar bir görsel analog skala kullanmış olabiliriz. Tanım olarak bu, sıralı kategorik bir değişken olmakla birlikte çeşitli gruplar arasında ağrı şiddetini karşılaştırmak için ortalamaları almamızda bir sakınca yoktur.

Diğer taraftan, değişkenler SPSS tarafından ordinal, nominal ve numerik olarak ayırt edilir (tam sayılı ve sürekli numerik veriler aynı şekilde işlem görür).



Veri çeşitlerini ayırt etmedeki önemli bir nokta da veri toplama sırasında karşımıza çıkmaktadır. Değişkenimizin numerik olması halinde veriyi de numerik olarak toplamalıyız. Sınıflandırma yapmak istersek daha sonra kategorilere ayırabiliriz. Örneğin, yaş değişkenini sorgulayan bir ankette açık uçlu bir şekilde

*“Kaç yaşındasınız?:\_\_\_\_\_\_” sorulmalıdır.*

*Sorumuz “Kaç yaşındasınız? a) 20’den az, b) 20-40, c) 41-60 d) 60’dan fazla”*

şeklinde olmamalıdır.

## Türetilmiş veriler

Tıpta yüzde, oran, orantı ve hız gibi türetilmiş verilerle de sıkça karşılaşırız.

Hastalarımızın durumlarındaki değişimleri takip etmek için bulgularımızı genellikle **yüzde** olarak ifade ederiz. Örneğin

*“Astımlı hastamızın FEV1 değerlerinde tedavi sonrasında %30 iyileşme oldu”*

deriz.

Bir bütünün iki parçasının birbirine bölünmesiyle **oran** (ratio) elde edilir. Örneğin,

*“Araştırmamızdaki bayan/erkek oranı 225/212 idi”.*

Bütünün parçalarından birinin bütüne bölünmesine ise **orantı** (proportion) denilir. Örneğin,

*“Araştırmamıza katılanların 225/437’si bayandı”. Orantıyı yüzde olarak ta ifade edebiliriz. Bu durumda “Araştırmamıza katılanların %51,5’i (225/437) bayandı.”*

deriz.

Birey sayıları zamanla ilişkili olarak ifade edilmesi durumunda **hız**’dan (rate) bahsederiz. Hız, epidemiyolojik çalışmalarda sıkça kullanılan bir terimdir.

*Neonatal ölüm hızı* (bir yıl içerisinde 0-28 günlükken ölen bebek sayısı/ bir yıldaki toplam doğum sayısı), *kaba doğum hızı* (her 1.000 kişilik nüfus başına yıllık doğum sayısı)

hız terimine örnek olarak verilebilir. Diğer taraftan dilimizde “hız” olarak ifade edilen *anne ölüm hızı* (belli bir dönemde meydana gelen anne ölümlerinin aynı dönemdeki canlı doğumlara bölünmesi) aslında bir orandır.

## Durdurulmuş (censored) veri

Bazen verilerimizin kesin durumu hakkında bilgimiz olmayabilir. Bu durumda durdurulmuş (censored, gizli) veriden bahsedebiliriz. Verilerimizin durdurulmuş olması istatistik açıdan istenmeyen bir durumdur. Mümkünse verinin gerçek değerini elde etmeliyiz veya sonuçlarımızı yazarken bu durumu belirtmeliyiz. İki şekilde durdurulmuş veri söz konusu olabilir:

* Ölçümle elde edeceğimiz veriler için ölçüm aracımızın belli bir değerin üzerini veya aşağısını saptayamaması durumunda. Laboratuvarımızın sonuçları bildirirken aşırı yüksek ve aşırı düşük değerleri “hi” veya “low” olarak ifade edip gerçek değerini veremediği dikkatinizi çekmiştir. Parmaktan kolesterol ölçümü yapan cihazların genelde 70 mg/dl ve altındaki değerleri okuyamayıp “low” olarak belirtmesini örnek olarak verebiliriz.
* Bazen de araştırmamız tamamlanmadan çeşitli nedenlerle (ölüm, yer değiştirme, araştırmaya katılmamaya karar verme...) hastalarımız ayrılabilirler. Bu durumda ölçmek istediğimiz değişkenin son durumu hakkında veri elde edemeyebiliriz. Örneğin, iki farklı girişimin kilo vermedeki etkisini araştırdığımızı ve hastalarımızın her ay kilolarını ölçtüğümüzü düşünelim. Belli bir aşamada hastanın ayrılması durumunda araştırmanın sonlanma noktasındaki veriyi elde edemeyeceğiz.