**Introductory Statistics for Data Science**

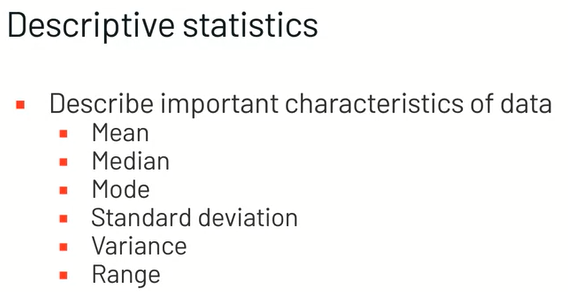
**Learning Objectives**

* Use hypothesis testing to answer a series of questions about the data set [Veri seti hakkında bir dizi soruyu cevaplamak için hipotez testini kullanın]
* Describe how outlier detection techniques can be used to determine if a data point is an outlier given a specific distribution [Bir veri noktasının belirli bir dağılım verilen bir aykırı değer olup olmadığını belirlemek için aykırı değer tespit tekniklerinin nasıl kullanılabileceğini açıklayın]
* Demonstrate the use of hypothesis tests with other hypotheses and other probability distributions [Diğer hipotezler ve diğer olasılık dağılımları ile hipotez testlerinin kullanımını göstermek]
* Demonstrate an understanding of hypothesis testing concepts [Hipotez testi kavramlarını anladığını gösterme]
* Demonstrate the use of a hypothesis test to answer a real-world question using normally-distributed data [Normal olarak dağıtılmış verileri kullanarak gerçek dünyadaki bir soruyu yanıtlamak için bir hipotez testinin kullanımını gösterin]
* Describe hypothesis testing as a tool for connecting inferential statistics, probability and the scientific method [Çıkarımsal istatistik, olasılık ve bilimsel yöntemi birbirine bağlamak için bir araç olarak hipotez testini tanımlayın]
* Identify one real world scenario that could be modeled by a discrete probability distribution and another real world scenario that can be modeled by a continuous probability distribution [Kesintili bir olasılık dağılımı ile modellenebilecek bir gerçek dünya senaryosu ve sürekli bir olasılık dağılımı ile modellenebilecek başka bir gerçek dünya senaryosu tanımlayın.]
* Apply hypothesis testing and outlier techniques to learn from data [Verilerden öğrenmek için hipotez testi ve aykırı değer tekniklerini uygulayın]
* Describe a few common continuous probability distributions and demonstrate how to use them practically to model data [Birkaç yaygın sürekli olasılık dağılımını tanımlayın ve bunların verileri modellemek için pratik olarak nasıl kullanılacağını gösterin]
* Use a Bernoulli distribution to model a real-world scenario [Gerçek dünya senaryosunu modellemek için bir Bernoulli dağılımı kullanın]
* Describe a few common discrete probability distributions [Birkaç yaygın ayrık olasılık dağılımını tanımlayın]
* Demonstrate the practical use of discrete probability distributions to model data [Verileri modellemek için ayrık olasılık dağılımlarının pratik kullanımını gösterin]
* Demonstrate an understanding of probability distributions and how they represent data [Olasılık dağılımlarını ve bunların verileri nasıl temsil ettiğini anladığını gösterme]
* Describe the concept behind probability distributions how they can represent data and random variables [Olasılık dağılımlarının ardındaki kavramı, verileri ve rastgele değişkenleri nasıl temsil edebileceklerini açıklayın]
* Recall foundational concepts about statistics presented in this course [Bu derste sunulan istatistikle ilgili temel kavramları hatırlayın]
* Explain how discrete and continuous probability distributions can be used to model data. [Ayrık ve sürekli olasılık dağılımlarının verileri modellemek için nasıl kullanılabileceğini açıklayın.]
* Compare and contrast discrete and continuous probability distributions [Ayrık ve sürekli olasılık dağılımlarını karşılaştırın ve karşılaştırın]
* Use the basic rules of discrete probability to complete a series of exercises [Bir dizi alıştırmayı tamamlamak için temel ayrık olasılık kurallarını kullanın.]
* Describe the basic rules of discrete probability [Ayrık olasılığın temel kurallarını tanımlayın]
* Compare and contrast descriptive statistics and inferential statistics [Tanımlayıcı istatistikleri ve çıkarımsal istatistikleri karşılaştırın ve karşılaştırın]
* Describe the field of probability and its relationship to inferential statistics [Olasılık alanını ve çıkarımsal istatistiklerle ilişkisini tanımlayın]
* Describe inferential statistics [Çıkarımsal istatistikleri tanımlayın]
* Use descriptive statistics to gather information about a data set [Bir veri seti hakkında bilgi toplamak için tanımlayıcı istatistikleri kullanın]
* Describe descriptive statistics [Tanımlayıcı istatistikleri tanımlayın]
* Describe statistics as a tool for learning from data [İstatistikleri verilerden öğrenmek için bir araç olarak tanımlayın]
* Apply foundational data science concepts from inferential statistics [Çıkarımsal istatistiklerden temel veri bilimi kavramlarını uygulayın]

# **Probability Distributions**

# **Statistics and Probability Review**

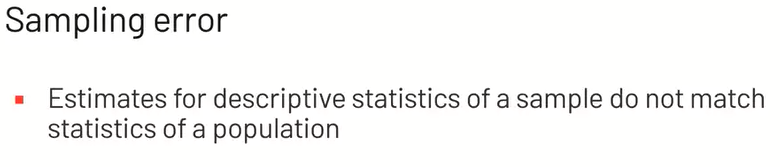
Hello, and welcome back. [Merhaba ve tekrar hoş geldiniz.] In this lesson, we will learn about probability distributions and how they can be used to model data. [Bu derste, olasılık dağılımlarını ve bunların verileri modellemek için nasıl kullanılabileceğini öğreneceğiz.] But first, let's recap what we've covered in a previous lesson. [Ama önce, bir önceki derste ele aldıklarımızı tekrarlayalım.]



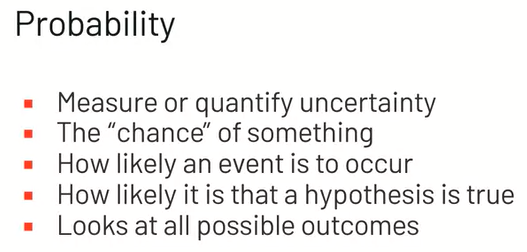
Remember that we can take a sample, or subset of data from some population, and calculate descriptive statistics to describe that sample. [Bir popülasyondan bir örnek veya veri alt kümesi alabileceğimizi ve bu örneği tanımlamak için tanımlayıcı istatistikleri hesaplayabileceğimizi unutmayın.] These include measures of central tendency, such as mean, median, and mode, and measures of dispersion such as standard deviation, variance, and range. [Bunlar, ortalama, medyan ve mod gibi merkezi eğilim ölçülerini ve standart sapma, varyans ve aralık gibi dağılım ölçülerini içerir.]



But what we usually want to do is use the sample data to make generalizations or inferences about some larger population. [Ancak genellikle yapmak istediğimiz şey, daha büyük bir popülasyon hakkında genellemeler veya çıkarımlar yapmak için örnek verileri kullanmaktır.] We're able to do this using inferential statistics. [Bunu çıkarımsal istatistikler kullanarak yapabiliriz.] In order to do so, however, we need to be confident that the sample accurately represents the population as a whole, and as unbiased as far as possible. [Ancak bunu yapabilmek için, örneğin popülasyonu bir bütün olarak doğru bir şekilde temsil ettiğinden ve mümkün olduğunca tarafsız olduğundan emin olmamız gerekir.]

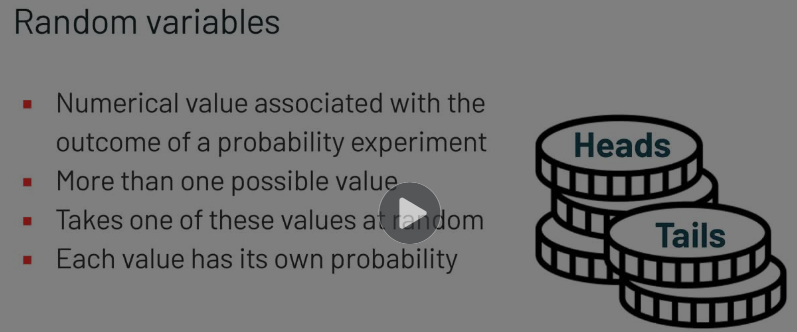


Even so, we know that there will be some sampling error, meaning that our estimates for the descriptive statistics of the sample won't perfectly match the statistics of the population. [Öyle olsa bile, bir miktar örnekleme hatası olacağını biliyoruz; bu, örneğin tanımlayıcı istatistiklerine ilişkin tahminlerimizin, popülasyonun istatistikleriyle tam olarak eşleşmeyeceği anlamına gelir.] This discrepancy between the sample and the population is what we refer to as sampling error. [Örneklem ve popülasyon arasındaki bu tutarsızlık, örnekleme hatası olarak adlandırdığımız şeydir.] Inferential statistics incorporates estimates of this error in their calculations, so that we can determine whether the patterns and associations seen in our sample actually generalize to the population. [Çıkarımsal istatistikler, bu hatanın tahminlerini hesaplamalarına dahil eder, böylece örneğimizde görülen kalıpların ve ilişkilerin gerçekten popülasyona genellenip genellenmediğini belirleyebiliriz.]

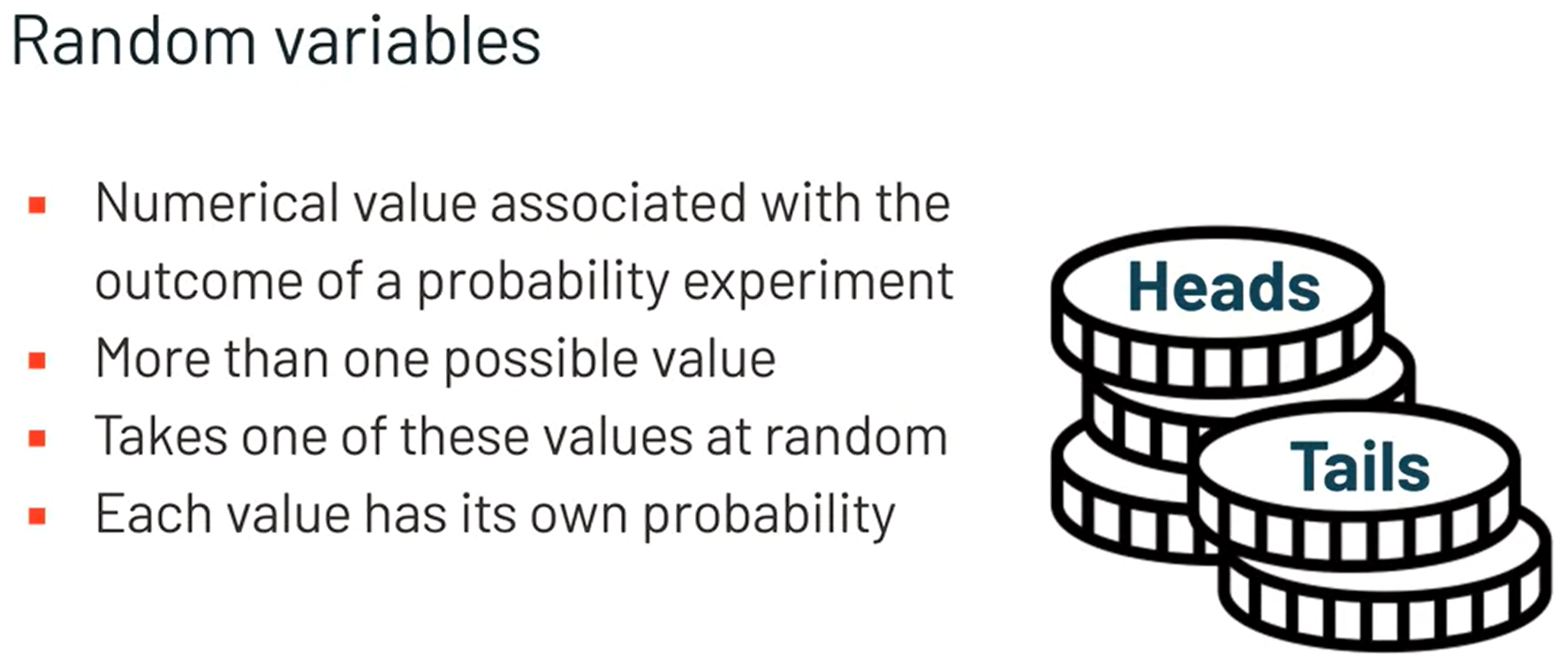


Finally, in the last lesson, we learned the basic rules of probability with discrete variables. [Son derste, kesikli değişkenlerle olasılığın temel kurallarını öğrendik.] Now, that we've reviewed the previous lesson, we'll talk about probability distributions and how they can represent data and random variables. [Şimdi, önceki dersi gözden geçirdiğimize göre, olasılık dağılımları ve bunların verileri ve rastgele değişkenleri nasıl temsil edebileceği hakkında konuşacağız.]

# **An Introduction to Probability Distributions**

****

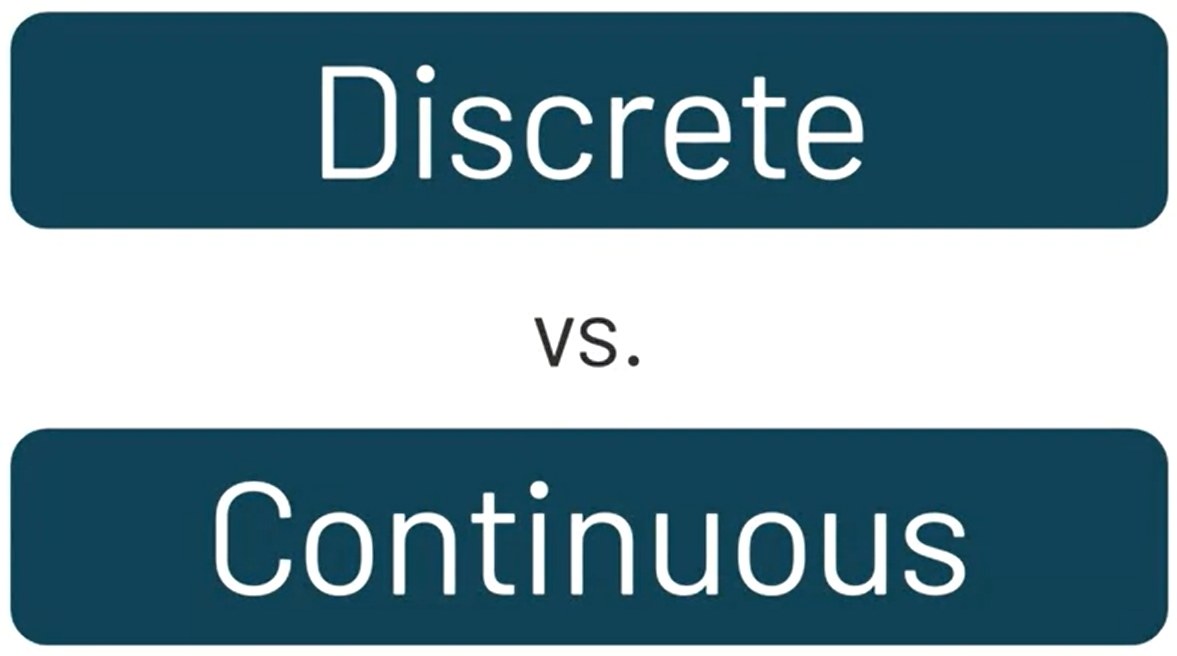
Probability distributions are a method used to quickly summarize the probabilities associated with the values of a random variable. [Olasılık dağılımları, rastgele bir değişkenin değerleriyle ilişkili olasılıkları hızlı bir şekilde özetlemek için kullanılan bir yöntemdir.] In this video, we'll describe the concept behind probability distributions and how they can represent data and random variables. [Bu videoda, olasılık dağılımlarının ardındaki kavramı ve bunların verileri ve rastgele değişkenleri nasıl temsil edebileceğini açıklayacağız.]



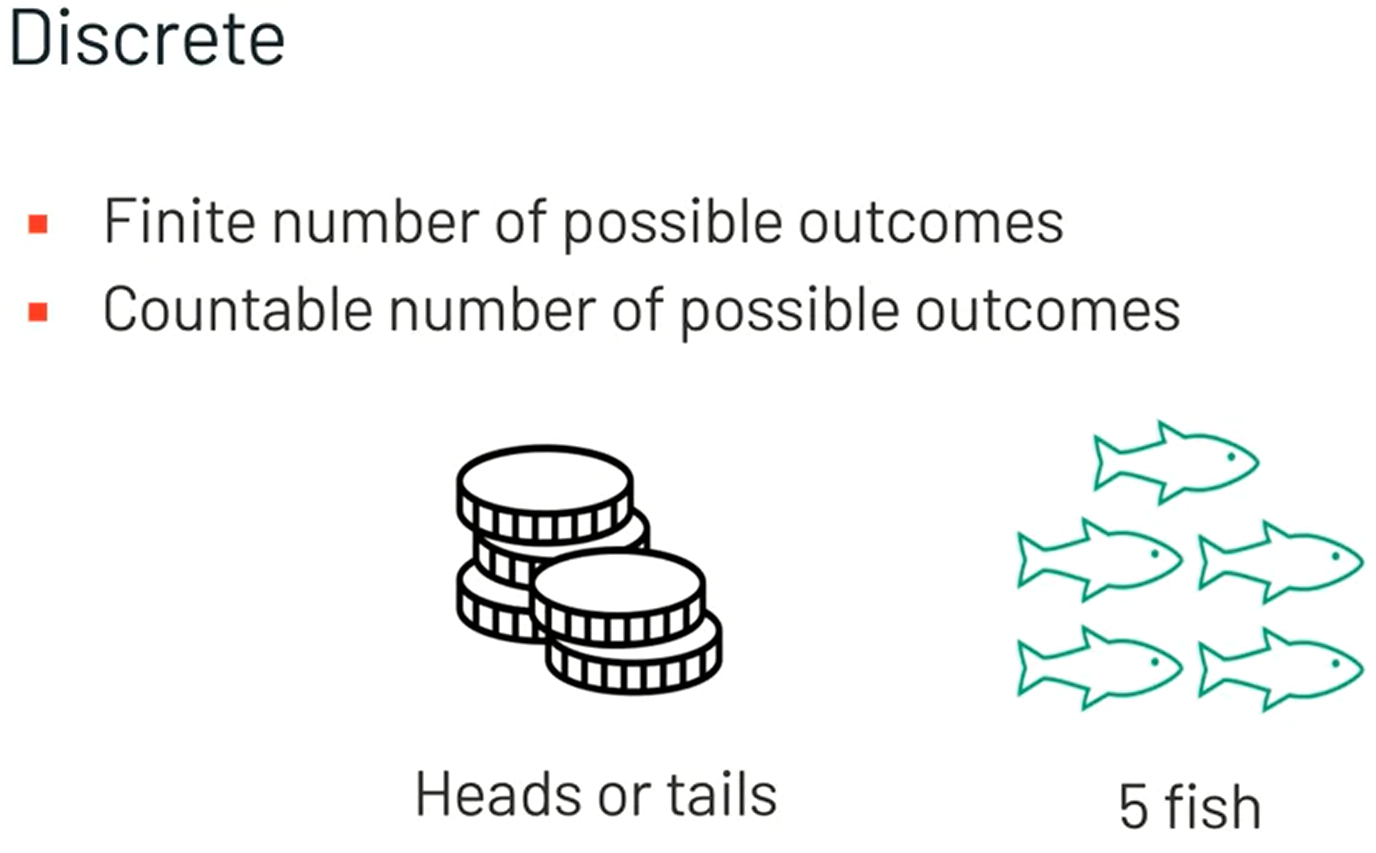
A random variable represents a numeric value associated with the outcome of a probability experiment. [Rastgele bir değişken, bir olasılık deneyinin sonucuyla ilişkili sayısal bir değeri temsil eder.] A random variable has more than one possible value, and it will take one of these values at random, each with a certain probability. [Rastgele bir değişkenin birden fazla olası değeri vardır ve bu değerlerden her biri belirli bir olasılıkla rastgele birini alacaktır.] For example, the side on which a coin lands is a random variable with two possible values, heads and tails, each with a probability of 0.5. [Örneğin, bir madalyonun düştüğü taraf, her biri 0,5 olasılıkla yazı ve tura olmak üzere iki olası değere sahip rastgele bir değişkendir.]



The sample space of a random variable is the set of all possible values for the variable. [Rastgele bir değişkenin örnek uzayı, değişken için tüm olası değerlerin kümesidir.] Going back to the dice rolling example, if we roll one die and there are six possible outcomes, 1, 2, 3, 4, 5, and 6, therefore, this is the sample space of the random variable. [Zar atma örneğine geri dönersek, bir zar atarsak ve 1, 2, 3, 4, 5 ve 6 olmak üzere altı olası sonuç varsa, bu nedenle, bu rastgele değişkenin örnek alanıdır.]



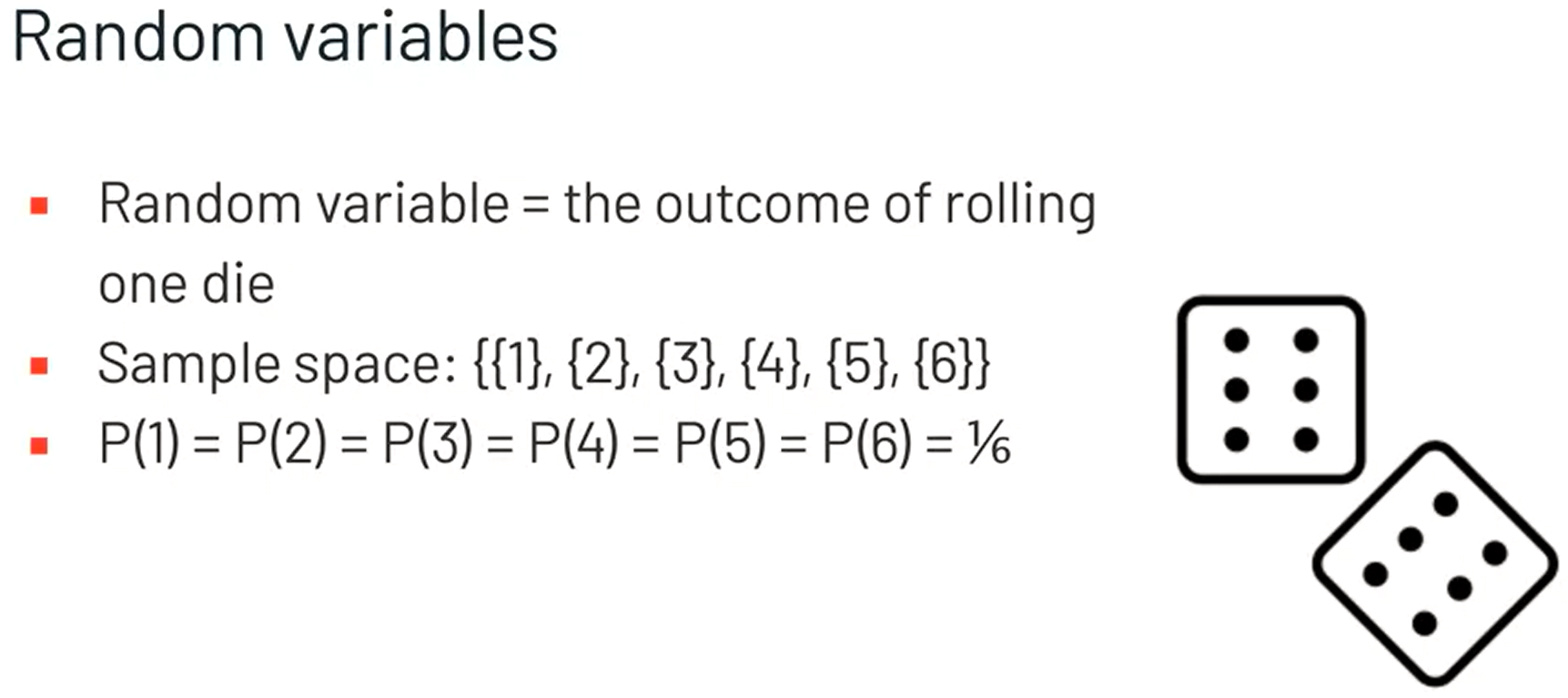
Random variables can be either discrete or continuous. [Rastgele değişkenler ayrık veya sürekli olabilir.] So far, we've only been talking about discrete random variables. [Şimdiye kadar sadece kesikli rastgele değişkenlerden bahsettik.]



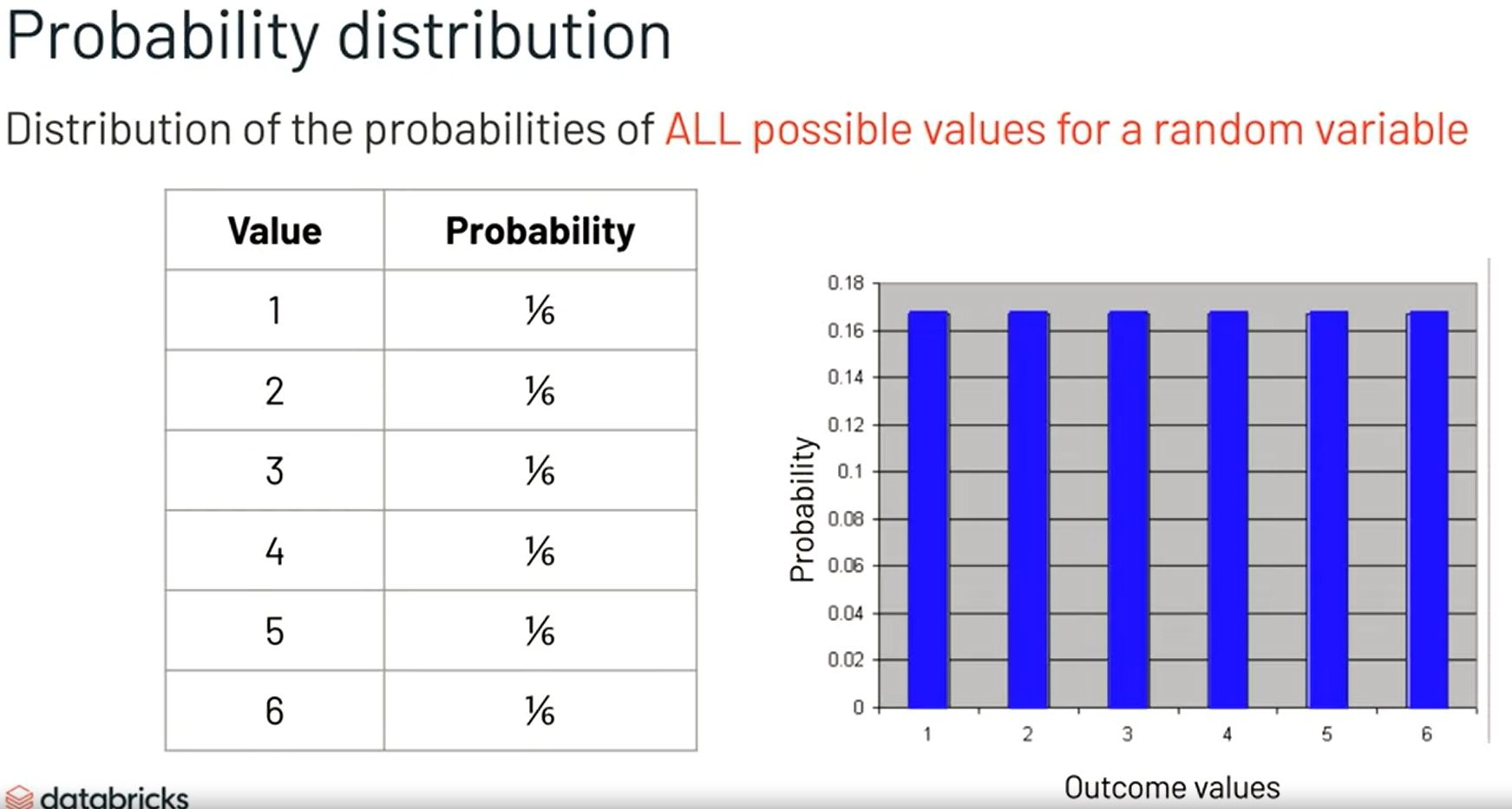
A discrete random variable has a finite countable number of possible outcomes. [Ayrık bir rastgele değişken, sonlu sayılabilir sayıda olası sonuca sahiptir.] As we talked about in the last lesson, these include things like flipping a coin or counting the number of fish. [Geçen derste bahsettiğimiz gibi, yazı tura atmak veya balık sayısını saymak gibi şeyler bunlara dahildir.] These outcomes can only be a whole number. [Bu sonuçlar sadece bir tam sayı olabilir.]



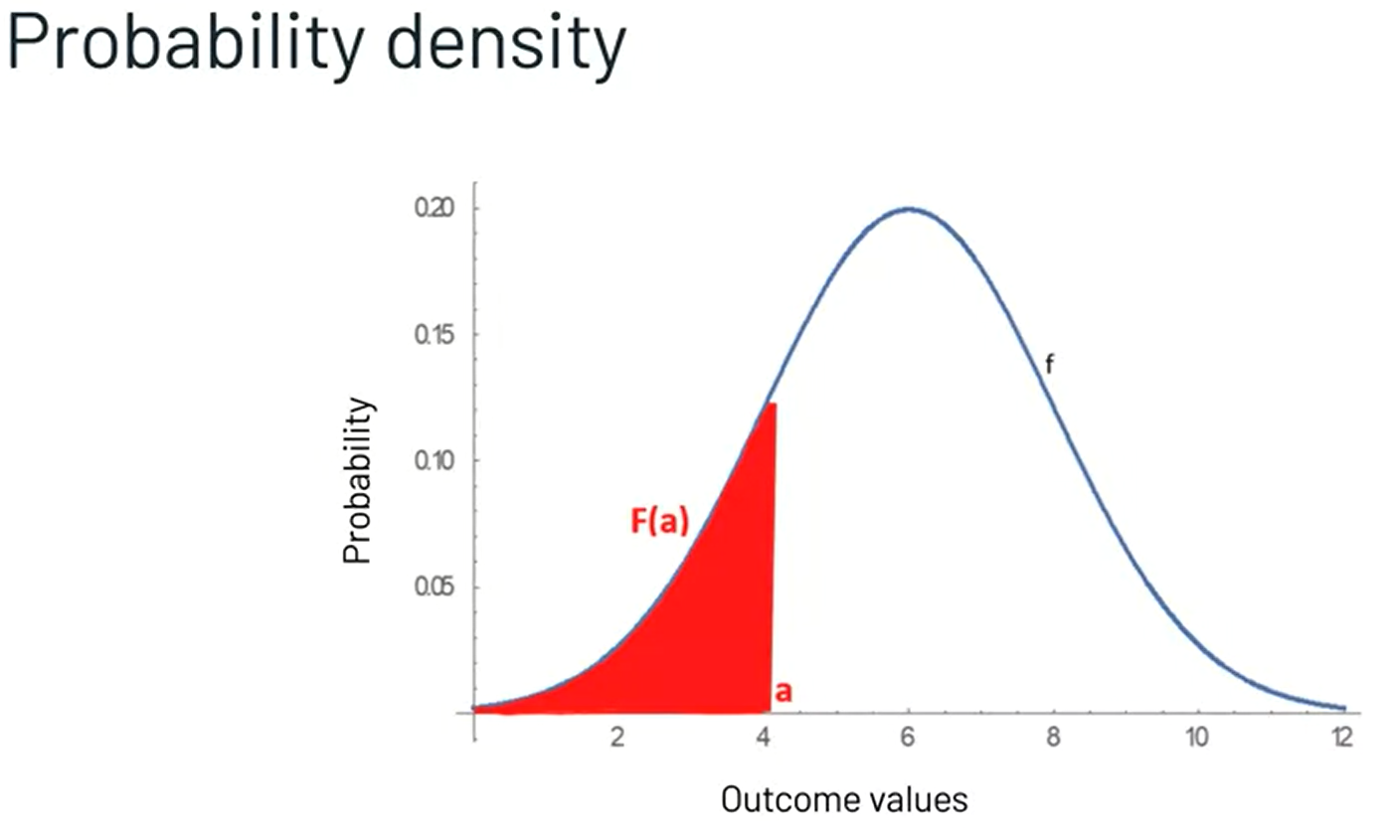
A continuous random variable has an infinite number of possible outcomes because it does not take the form of countable integers. [Sürekli bir rastgele değişken, sayılabilir tamsayılar biçimini almadığı için sonsuz sayıda olası sonuca sahiptir.] Think of a number line. [Bir sayı doğrusu düşünün.] The possible values along this line are not limited to whole numbers. [Bu satırdaki olası değerler tam sayılarla sınırlı değildir.] You can have a variable equal to the number 2, 2.1, or 2.0001, with infinite values in between. [Aralarında sonsuz değerler bulunan 2, 2.1 veya 20001 sayısına eşit bir değişkeniniz olabilir.] An example of a continuous random variable could be weight, or how long it takes to run a kilometer, or even the amount of sugar in a strawberry. [Sürekli rastgele değişkene bir örnek, ağırlık veya bir kilometreyi koşmanın ne kadar sürdüğü, hatta bir çileğin içindeki şeker miktarı olabilir.]



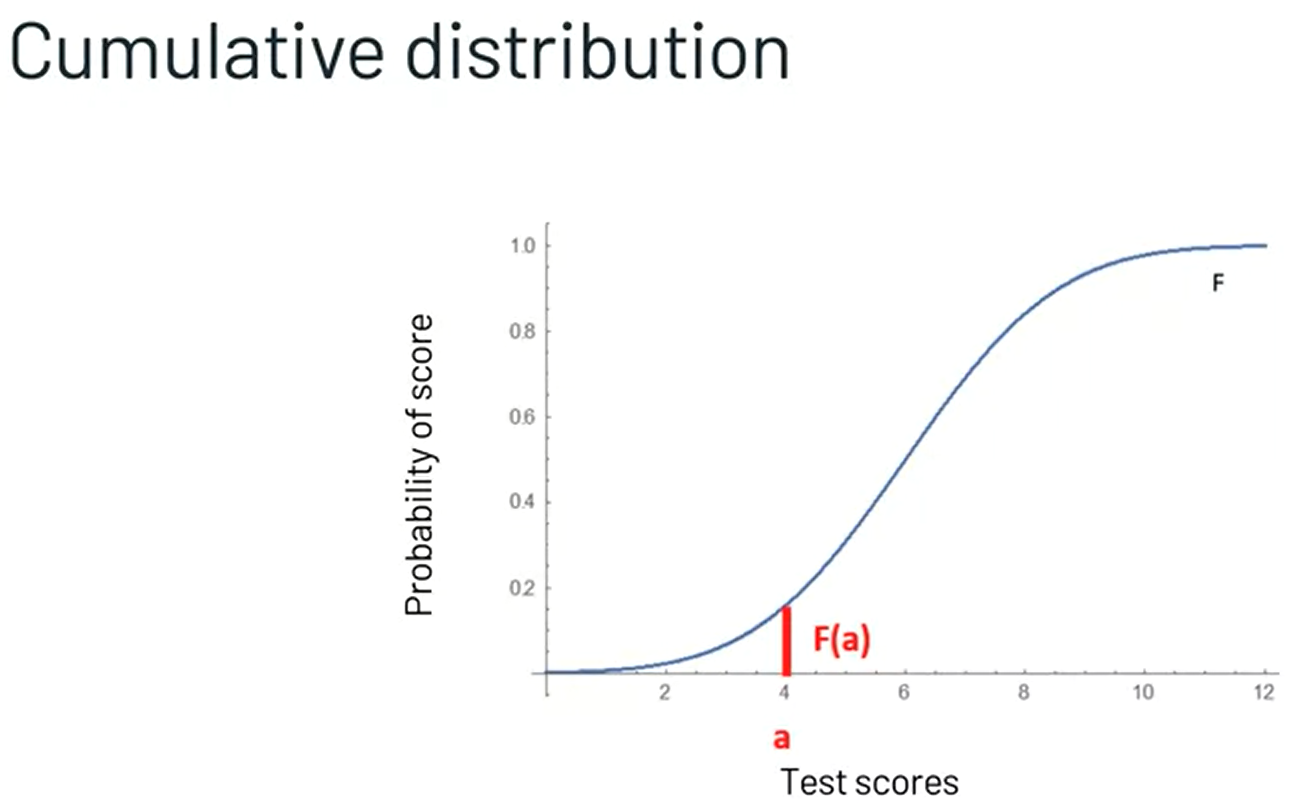
We're now going to make a jump from talking about single values and their probabilities to distributions of all of the possible values of a random variable. [Şimdi tek değerler ve bunların olasılıkları hakkında konuşmaktan bir rastgele değişkenin tüm olası değerlerinin dağılımlarına geçeceğiz.] Once we have the sample space for the random variable, we need to assign probabilities to each of the values in that sample space. [Rastgele değişken için örnek uzaya sahip olduğumuzda, o örnek uzaydaki değerlerin her birine olasılık atamamız gerekir.] For example, there are six outcomes that can occur when we roll a single die. [Örneğin, tek bir zar attığımızda ortaya çıkabilecek altı sonuç vardır.] This example is a little boring because each of the six possible values in the sample space for this experiment has the exact same probability of occurring. [Bu örnek biraz sıkıcı çünkü bu deney için örnek uzaydaki altı olası değerin her biri aynı gerçekleşme olasılığına sahip.]



Once we have the sample space defined, we can create a probability distribution. [Örnek uzayı tanımladığımızda, bir olasılık dağılımı oluşturabiliriz.] That is the distribution of the probabilities of all of the possible values for the random variable. [Bu, rastgele değişken için tüm olası değerlerin olasılıklarının dağılımıdır.] A probability distribution maps these values to their probabilities. [Bir olasılık dağılımı, bu değerleri olasılıklarına göre eşleştirir.] Probability distributions can be displayed as tables, like you see on the left, or more frequently as bar graph plots for discrete random variables. [Olasılık dağılımları, solda gördüğünüz gibi tablolar olarak veya daha sık olarak ayrık rastgele değişkenler için çubuk grafik çizimleri olarak görüntülenebilir.]



For continuous variables, we call their probability distribution a probability density. [Sürekli değişkenler için olasılık dağılımlarına olasılık yoğunluğu diyoruz.] Since we don't have discrete values where we can calculate an exact probability for each, this will end up looking more like a cloud or a density of values and probabilities. [Her biri için kesin bir olasılık hesaplayabileceğimiz ayrık değerlerimiz olmadığı için, bu daha çok bir buluta veya bir değerler ve olasılık yoğunluğuna benzeyecektir.]



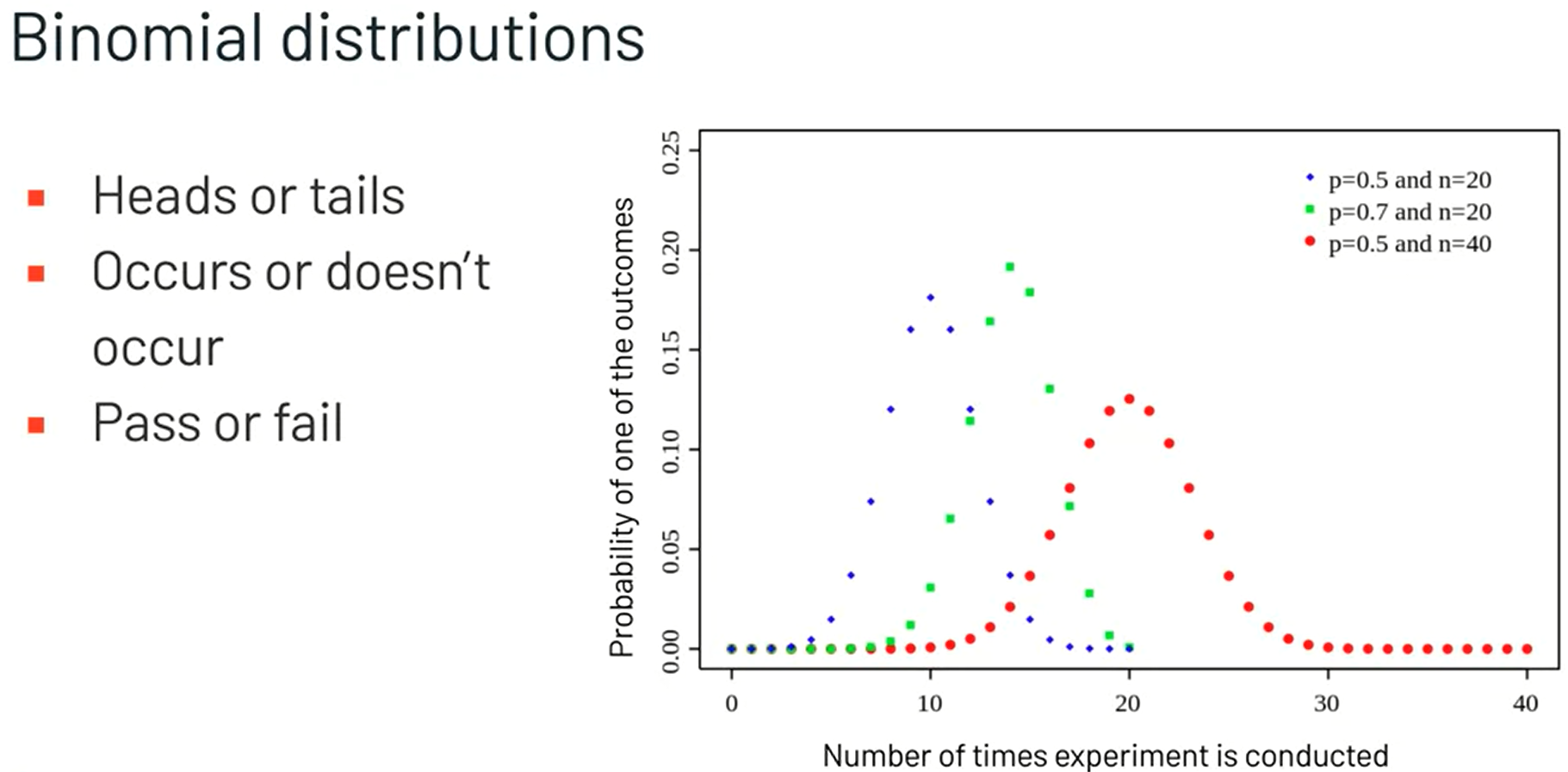
There are also cumulative probability distributions which map all of the possible outcomes less than or equal to a specified value to a single probability. [Ayrıca, belirli bir değere eşit veya daha az olası tüm sonuçları tek bir olasılığa eşleyen kümülatif olasılık dağılımları da vardır.] This gives us the probability that a value lies within a certain range. [Bu bize bir değerin belirli bir aralıkta olma olasılığını verir.] For example, we might want to know the probability of students in a class scoring above a 70 percent on a test, which we can also describe as scoring between 70 and 100 percent. [Örneğin, bir sınıftaki öğrencilerin bir testte yüzde 70'in üzerinde puan alma olasılığını bilmek isteyebiliriz, bunu yüzde 70 ile yüzde 100 arasında puanlama olarak da tanımlayabiliriz.]



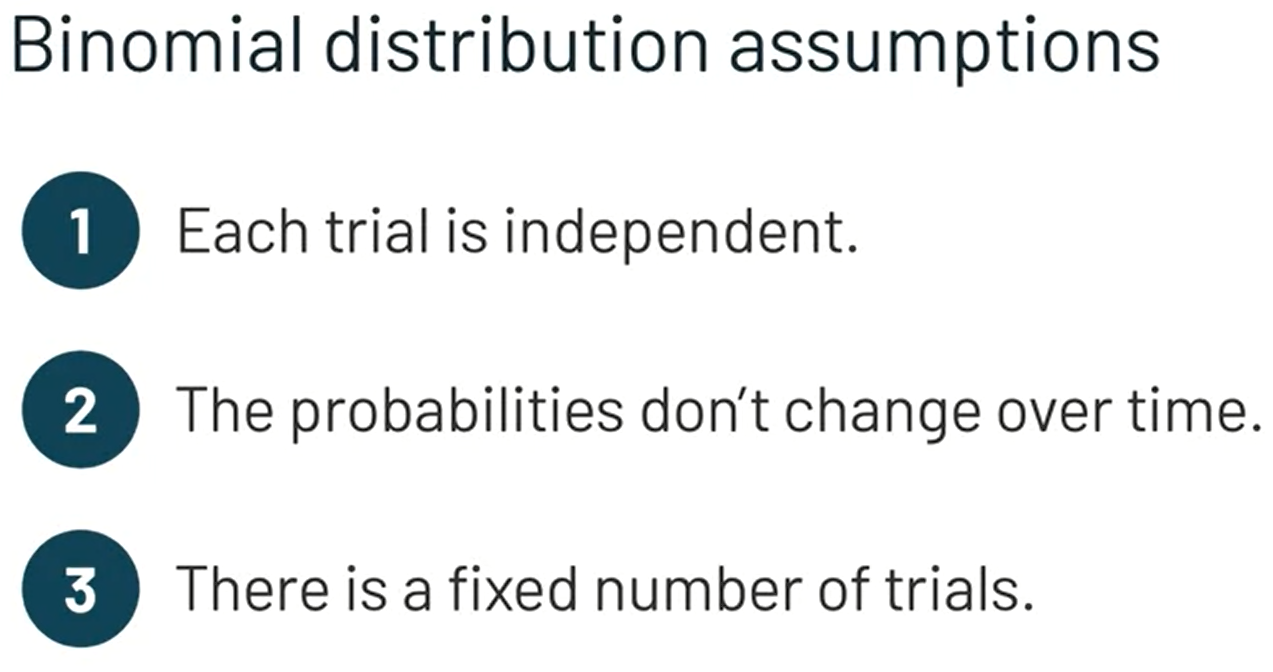
Each probability distribution has some general properties that define the distribution. [Her olasılık dağılımı, dağılımı tanımlayan bazı genel özelliklere sahiptir.] Mathematically, these properties are referred to as moments and include the expected value, variance, skewness, and kurtosis. [Matematiksel olarak, bu özellikler momentler olarak adlandırılır ve beklenen değeri, varyansı, çarpıklığı ve basıklığı içerir.] These properties all describe a different quality of the data set and are important for making inferences. [Bu özelliklerin tümü, veri kümesinin farklı bir kalitesini tanımlar ve çıkarım yapmak için önemlidir.] There are many types of probability distributions, and they vary based on the properties of the random variables in the sample data. [Pek çok olasılık dağılımı vardır ve bunlar örnek verilerdeki rastgele değişkenlerin özelliklerine göre değişir.] You will now answer some questions to check your understanding of probability distributions and then in the following lessons, we'll go into more detail on discrete and continuous distributions and their properties. [Şimdi olasılık dağılımları konusundaki anlayışınızı kontrol etmek için bazı soruları yanıtlayacaksınız ve sonraki derslerde ayrık ve sürekli dağılımlar ve özellikleri hakkında daha fazla ayrıntıya gireceğiz.]

# **Discrete Probability Distributions**

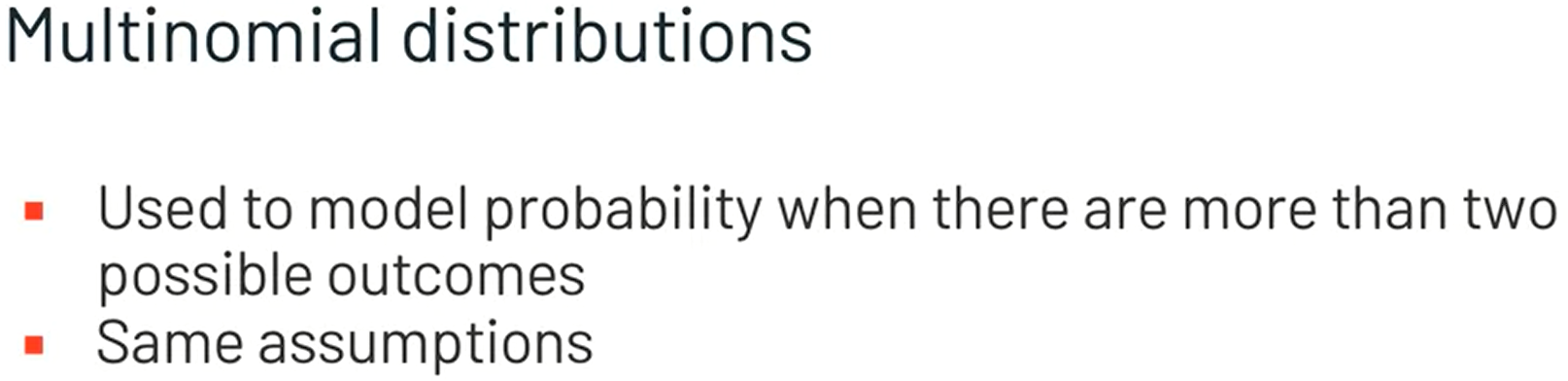
Remember from the previous lesson, that probability distributions are defined and distinguished by their properties like variance and skieunice. [Bir önceki dersten hatırlayın, olasılık dağılımları varyans ve skyeunice gibi özellikleriyle tanımlanır ve ayırt edilir.] We call these moments. [Bu anlar diyoruz.] In our previous example, we talked about rolling a single die. [Bir önceki örneğimizde tek bir zar atmaktan bahsetmiştik.] Since the probability of each possible outcome for this experiment is the same, it has a uniform distribution. [Bu deney için olası her sonucun olasılığı aynı olduğundan, düzgün bir dağılıma sahiptir.] We'll now describe some of the other common types of discrete probability distributions. [Şimdi ayrık olasılık dağılımlarının diğer yaygın türlerinden bazılarını tanımlayacağız.]



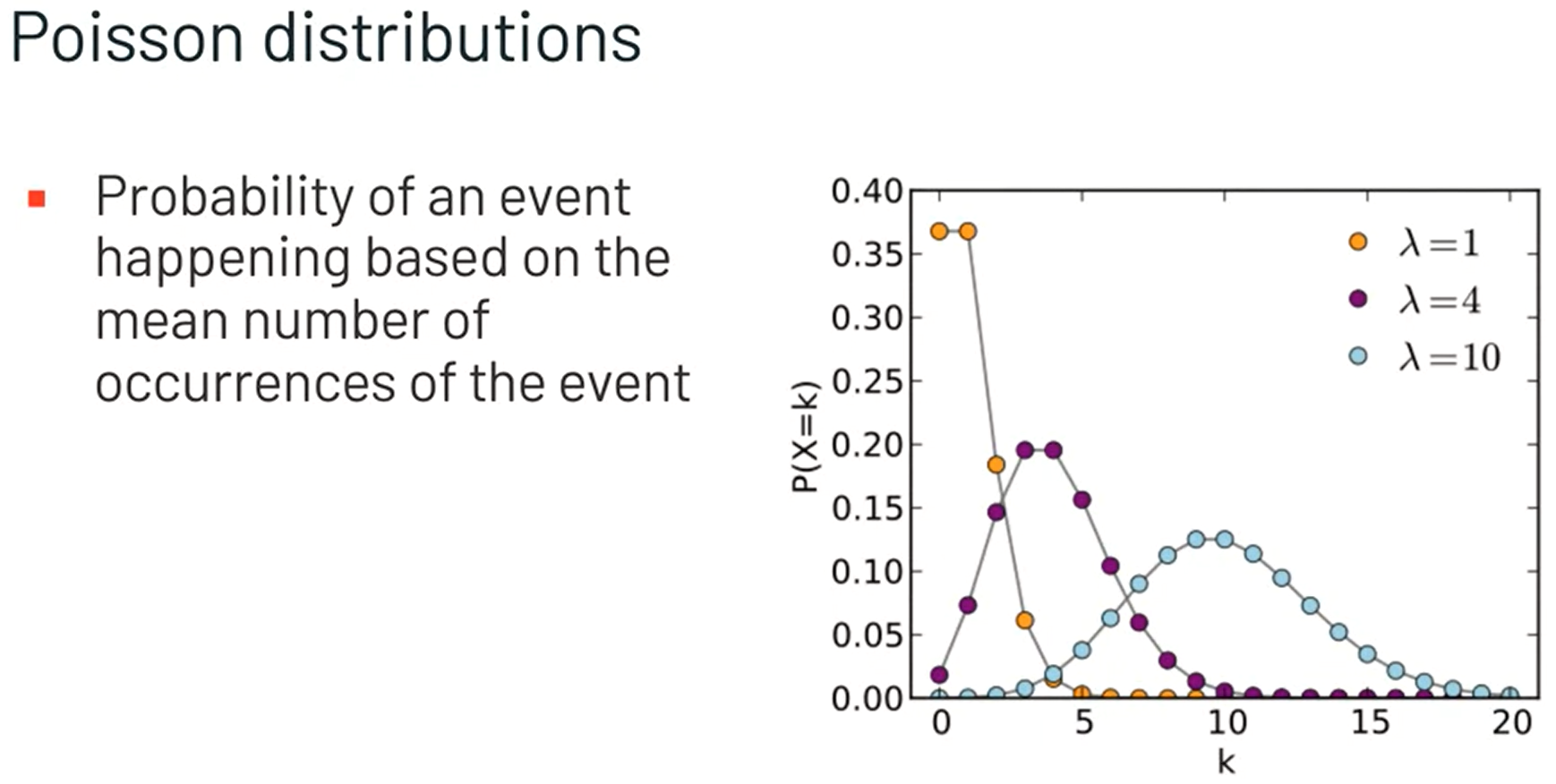
A binomial probability distribution models, binary outcomes like a coin toss where there are only two possible outcomes. [İki terimli olasılık dağılım modelleri, yalnızca iki olası sonucun olduğu bir yazı tura gibi ikili sonuçlar.] Heads or tails, for example, or an event occurring or not occurring or a pass or fail testing scenario. [Örneğin yazılar veya turalar veya meydana gelen veya gelmeyen bir olay veya başarılı veya başarısız bir test senaryosu.] Basically anything you can think of that can be framed as a success, a thing occurs or a failure, a thing does not occur can be represented as a binomial distribution. [Temel olarak, bir başarı, bir şey meydana gelir veya bir başarısızlık, bir şey olmaz olarak çerçevelenebilecek herhangi bir şey, bir binom dağılımı olarak temsil edilebilir.]



In addition to only two possible values for the outcome, the binomial distribution makes a few other assumptions about your data. [Sonuç için yalnızca iki olası değere ek olarak, binom dağılımı verileriniz hakkında birkaç varsayım daha yapar.] It assumes that each trial is independent, meaning the probabilities of one trial don't affect the probabilities of the next trial. [Her denemenin bağımsız olduğunu varsayar, yani bir denemenin olasılıkları bir sonraki denemenin olasılıklarını etkilemez.] It also assumes the probability for two outcomes don't change over time. [Ayrıca, iki sonucun zaman içinde değişmeme olasılığını da varsayar.] The coin will not become more likely to land heads up next week than it is this week. [Madeni paranın gelecek hafta gelme olasılığı bu hafta olduğundan daha fazla olmayacak.] And finally, assumes there's a fixed number of trials that the distribution models. [Ve son olarak, dağıtım modellerinin sabit sayıda denemesi olduğunu varsayar.] I will toss the coin 10 times and the distribution models the probability of the two values for those 10 trials. [Madeni parayı 10 kez atacağım ve dağılım bu 10 deneme için iki değerin olasılığını modelleyecektir.] If the data meets those criteria, we can use a binomial distribution to model the probabilities. [Veriler bu kriterleri karşılıyorsa, olasılıkları modellemek için bir binom dağılımı kullanabiliriz.]



Multinomial distributions are similar to binomial distributions, but they're used to model probability when there are more than two possible outcomes for a variable or an experiment. [Çok terimli dağılımlar, iki terimli dağılımlara benzer, ancak bir değişken veya deney için ikiden fazla olası sonuç olduğunda olasılığı modellemek için kullanılırlar.] The other three criteria are the same however. [Ancak diğer üç kriter aynıdır.] Each trial is independent. [Her deneme bağımsızdır.] The probability doesn't change over time and there's a fixed number of trials. [Olasılık zamanla değişmez ve sabit sayıda deneme vardır.]



The Poisson distribution is used to display the probability of an event happening based on the mean number of occurrences of the event. [Poisson dağılımı, olayın ortalama gerçekleşme sayısına bağlı olarak bir olayın olma olasılığını göstermek için kullanılır.] Unlike a binomial experiment that has a set number of trials, a poisson experiment focuses only on the number of discrete occurrences over some interval. [Belirli sayıda denemeye sahip bir binom deneyinin aksine, bir poisson deneyi yalnızca belirli bir aralıktaki ayrık oluşumların sayısına odaklanır.] Some examples are stores that know they sell on average a certain number of an item and they can calculate the probability of selling more or less than that number. [Bazı örnekler, ortalama olarak belirli sayıda ürün sattıklarını bilen ve bu sayıdan daha fazla veya daha az satış yapma olasılığını hesaplayabilen mağazalardır.] Or we know on average our little fishes are just eight cm long. [Veya ortalama olarak küçük balıklarımızın sadece sekiz cm uzunluğunda olduğunu biliyoruz.] So we can determine the probability of fish as being four cm long or nine cm long. [Böylece balığın dört cm uzunluğunda veya dokuz cm uzunluğunda olma olasılığını belirleyebiliriz.] The events must be independent of one another. [Olaylar birbirinden bağımsız olmalıdır.] Just like the binomial and multi normal distributions. [Tıpkı binom ve çoklu normal dağılımlar gibi.]



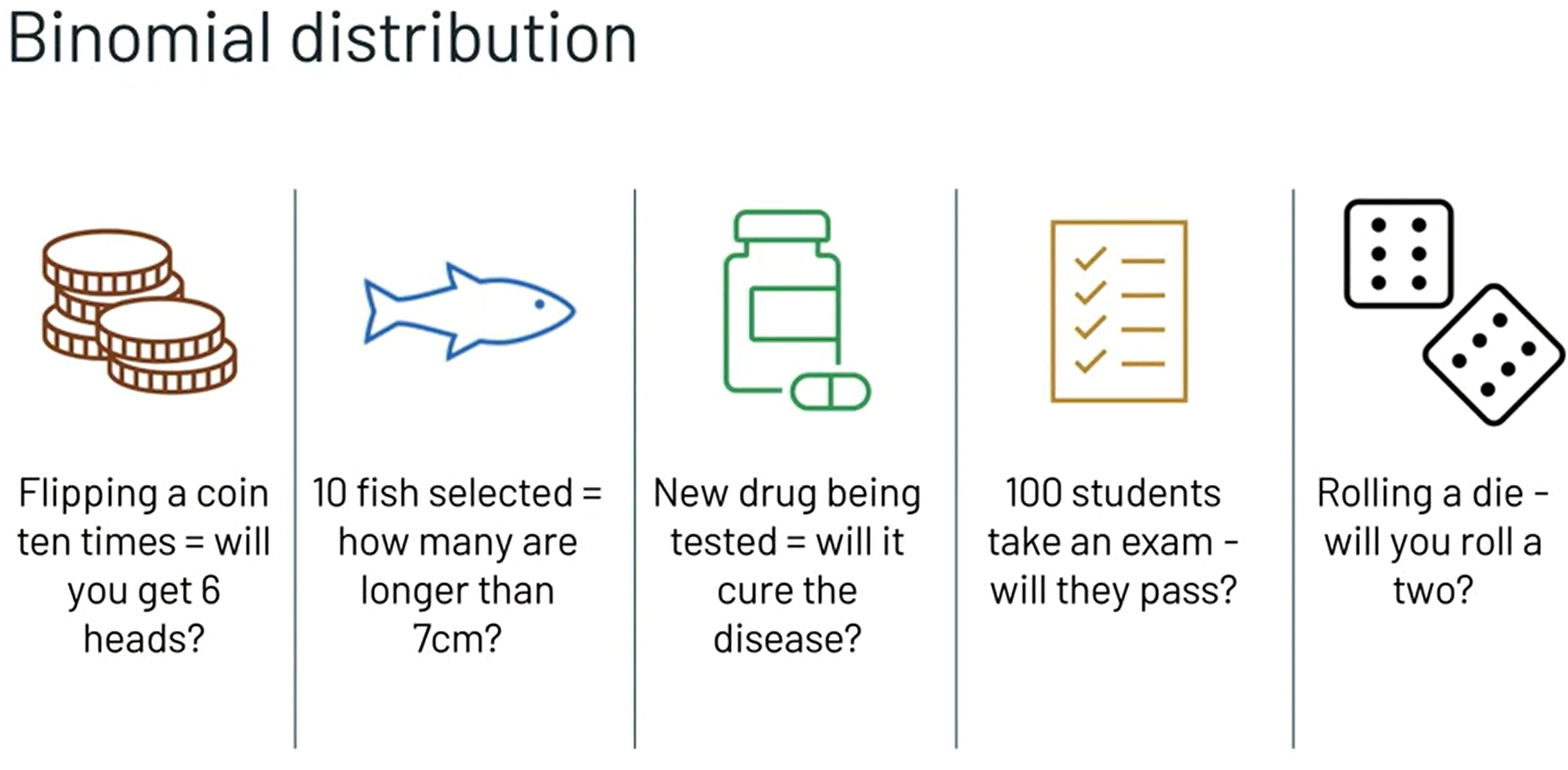
The Bernoulli distribution deals with bernoulli random variables, which are random variables that have only two possible outcomes. [Bernoulli dağılımı, yalnızca iki olası sonucu olan rastgele değişkenler olan bernoulli rastgele değişkenleriyle ilgilenir.] You might notice this sounds just like the binomial distribution and in fact, the two are closely related. [Bunun binom dağılımına benzediğini fark edebilirsiniz ve aslında ikisi yakından ilişkilidir.] All bernoulli distributions are binomial distributions, but the reverse is not true. [Tüm bernoulli dağılımları iki terimli dağılımlardır, ancak bunun tersi doğru değildir.] Not all binomial distributions are bernoulli distributions. [Tüm binom dağılımları bernoulli dağılımları değildir.] This is because a bernoulli distribution is a binomial distribution for a single trial of an experiment, such as flipping a coin, just one time or just one student taking a test once. [Bunun nedeni, bernoulli dağılımının bir denemenin tek bir denemesi için bir binom dağılımı olmasıdır, örneğin yazı tura atmak, sadece bir kez veya sadece bir öğrenci bir kez sınava girmek gibi.] This single random variable is also known as a bernoulli trial. [Bu tek rastgele değişken, bernoulli denemesi olarak da bilinir.]



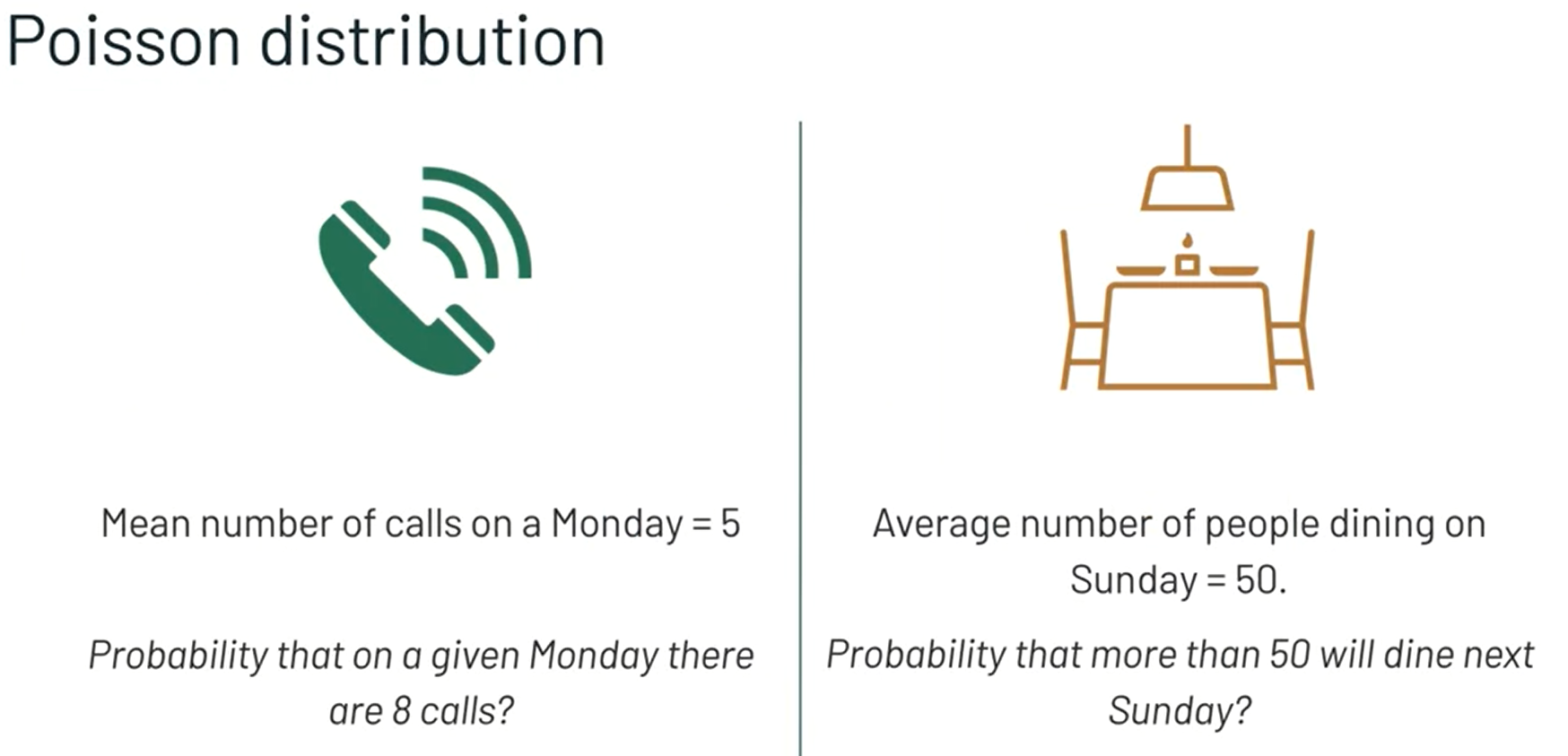
The two possible outcomes of bernoulli trials are usually referred to as a success or a failure. [Bernoulli denemelerinin iki olası sonucu genellikle başarı veya başarısızlık olarak adlandırılır.] These are labeled as, 1 for success and 0 for failure and their probabilities are p the probability of success and 1-p, or the complement of p as the probability of failure. [Bunlar başarı için 1 ve başarısızlık için 0 olarak etiketlenir ve olasılıkları p başarı olasılığı ve 1-p veya başarısızlık olasılığı olarak p'nin tamamlayıcısıdır.] A binomial distribution is the sum of independent and identically distributed bernoulli trials. [Binom dağılımı, bağımsız ve aynı şekilde dağılmış bernoulli denemelerinin toplamıdır.] Recall that this means, the outcome of each trial is unaffected by the outcome of the other trials and each trial has the same probability for the two outcomes. [Bunun, her denemenin sonucunun diğer denemelerin sonucundan etkilenmediği ve her denemenin iki sonuç için aynı olasılığa sahip olduğu anlamına geldiğini hatırlayın.] You now know about some of the most common types of discrete probability distributions. [Artık en yaygın kesikli olasılık dağılımlarından bazılarını biliyorsunuz.] Next we'll explore some of the practical use cases of discrete probability for modeling data. [Daha sonra, verileri modellemek için ayrık olasılığın bazı pratik kullanım durumlarını keşfedeceğiz.]

# **Discrete Probability Distributions Applications**

So far, we've been talking about the definitions of probability distributions in a pretty abstract way. [Buraya kadar olasılık dağılımlarının tanımlarından oldukça soyut bir şekilde bahsettik.] Now we'll look at some really world examples of when binomial and puts on distributions would be used to model data. [Şimdi, verileri modellemek için binom ve dağıtımların ne zaman kullanılacağına dair gerçekten dünya örneklerine bakacağız.]



For the binomial distribution one example is the probability of getting exactly six heads when you flip a coin 10 times. [Binom dağılımı için bir örnek, bir madeni parayı 10 kez çevirdiğinizde tam olarak altı tura gelme olasılığıdır.] In another example, if 10 fish are randomly selected, what is the probability of them being longer than seven centimeters? [Başka bir örnekte, rastgele 10 balık seçildiğinde, bunların yedi santimetreden uzun olma olasılığı nedir?] Exactly three of them being longer than seven centimeters? [Tam olarak üç tanesi yedi santimetreden uzun mu?] Keep in mind there either longer than seven centimeters or they're not. [Orada yedi santimetreden daha uzun olduğunu veya olmadığını unutmayın.] So this is binomial. [Yani bu iki terimli.] A third example, a new drug is being tested in eight clinical trials. [Üçüncü bir örnek, yeni bir ilaç sekiz klinik deneyde test ediliyor.] What's the probability that it cures the disease? [Hastalığı iyileştirme olasılığı nedir?] It either cures the disease. [Ya hastalığı tedavi eder.] It's successful or it doesn't cure the disease. [Başarılı olur ya da hastalığı tedavi etmez.] It's a failure. [Bu bir başarısızlık.] A different outcome is possible. [Farklı bir sonuç mümkündür.] The success or the failure for each of those eight clinical trials with our testing example when 100 students might take an exam. [100 öğrencinin bir sınava girebileceği test örneğimizle bu sekiz klinik araştırmanın her birinin başarısı veya başarısızlığı.] Remember that for each of those, 100 students with a binomial distribution were saying that each student has the same probability of passing that exam, and each student will either pass or they won't pass and finally heading back to our die. [Bunların her biri için binom dağılımına sahip 100 öğrencinin, her öğrencinin o sınavı geçme olasılığının aynı olduğunu ve her öğrencinin ya geçecek ya da geçemeyeceklerini söylediğini ve sonunda ölümümüze geri döndüğünü unutmayın.] Example to model the rolling of a die with a binomial distribution. [Binom dağılımına sahip bir kalıbın yuvarlanmasını modelleme örneği.] We need a binary outcome, right, like rolling a two. [İkili bir sonuca ihtiyacımız var, doğru, ikiyi yuvarlamak gibi.] So what would be the probability of us rolling exactly three twos when we roll a die 10 times? [Öyleyse, bir zarı 10 kez attığımızda tam olarak üç ikişer atma olasılığımız nedir?]



And now the Poisson distribution? [Ve şimdi Poisson dağılımı?] Suppose you knew that the mean number of calls to customer support on a Monday is five. [Pazartesi günü müşteri desteğine yapılan ortalama arama sayısının beş olduğunu bildiğinizi varsayalım.] What is the probability that on a given Monday, there would be eight calls or the average number of people dining in a restaurant for dinner on Sundays? [Belirli bir Pazartesi günü, Pazar günleri akşam yemeği için bir restoranda yemek yiyen ortalama kişi sayısı veya sekiz arama olma olasılığı nedir?] This 50? [Bu 50?] What's the probability that more than 50 people will die next Sunday, useful for sales and businesses to predict how much goods to order, arrange adequate staffing or determine for which hours and days it makes sense to be open. [Satışlar ve işletmeler için ne kadar mal sipariş edileceğini tahmin etmede, yeterli personel ayarlamada veya hangi saatler ve günlerde açık olmanın mantıklı olduğunu belirlemede faydalı olacak şekilde, önümüzdeki Pazar 50'den fazla kişinin ölme olasılığı nedir?] This can help companies improve their operational efficiency. [Bu, şirketlerin operasyonel verimliliklerini artırmalarına yardımcı olabilir.] Connecting these discrete probability distributions to the real world demonstrates their power, and the rest of the lesson will begin looking at probability distributions for continuous random variables. [Bu ayrık olasılık dağılımlarını gerçek dünyasına bağlamak, onların gücünü gösterir ve dersin geri kalanı sürekli rastgele değişkenler için olasılık dağılımlarına bakmaya başlayacaktır.]

# **Probability Distribution Lab Intro**

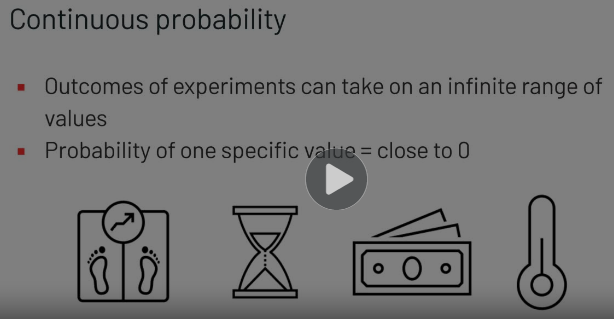
Now that you've learned all about discrete probability distributions, it's time for some hands on work. In the next lab, you will practice making some probability distributions. We hope you enjoy it.

# **Discrete Probability Lab**

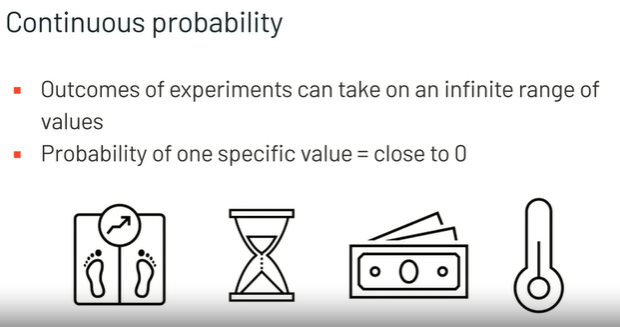
Next, you'll complete a lab on the Bernoulli distribution using Databricks.

To do so, log in to your Databricks Community Edition account and locate your DSFDA folder. Then, open Module 3, then click on "3.2.1 Lab - Bernoulli Distribution" to open your notebook.

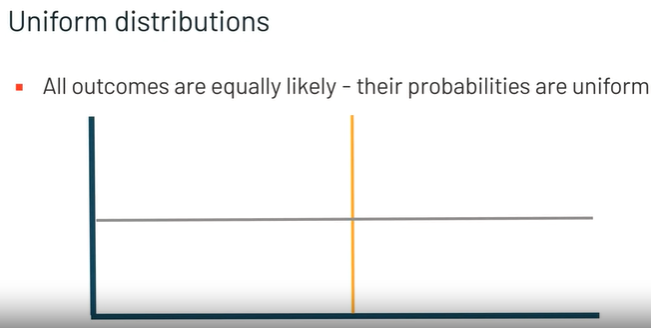
# **Continuous Probability Distributions**



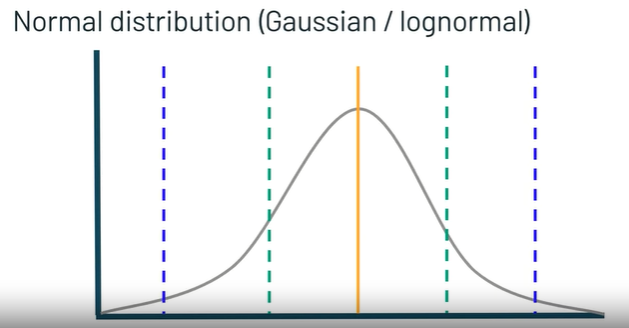
So far in this lesson, we've been talking exclusively about discrete probability distributions, where the outcomes of experiments are represented by countable whole numbers, like a dye that is either two or three, but not 2.134. [Bu derste şimdiye kadar, deney sonuçlarının sayılabilir tam sayılarla temsil edildiği, örneğin iki ya da üç olan ama 2.134 olmayan bir boya gibi, ayrık olasılık dağılımlarından bahsettik.]



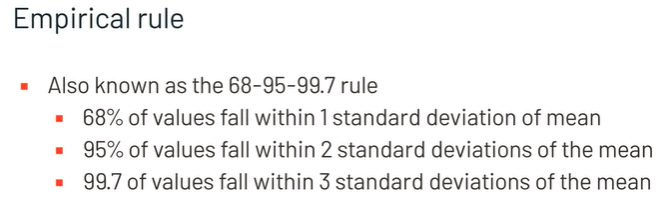
Continuous probability deals with scenarios where the outcomes of experiments can take on an infinite range of values. [Sürekli olasılık, deneylerin sonuçlarının sonsuz bir değer aralığı alabileceği senaryolarla ilgilenir.] Continuous random variables have values that occur along the scale, such as weight, lengths of time, amounts of money or temperature. [Sürekli rastgele değişkenler, ağırlık, zaman uzunlukları, para miktarları veya sıcaklık gibi ölçek boyunca meydana gelen değerlere sahiptir.] Because there are infinite possible values that the outcome can assume, the probability of it being anyone specific value is close to 0. [Sonucun alabileceği sonsuz olası değerler olduğundan, herhangi bir belirli değer olma olasılığı 0'a yakındır.] Therefore, we usually talk about the probability of an outcome being within a range of values rather than a specific value like we did with a coin toss or rolling a dye. [Bu nedenle, genellikle yazı tura veya boya yuvarlama ile yaptığımız gibi belirli bir değerden ziyade bir değer aralığı içinde bir sonucun olasılığı hakkında konuşuruz.] Continuous probability distributions are expressed with a formula known as a probability density function that describes the shape of the distribution. [Sürekli olasılık dağılımları, dağılımın şeklini tanımlayan olasılık yoğunluk fonksiyonu olarak bilinen bir formülle ifade edilir.]



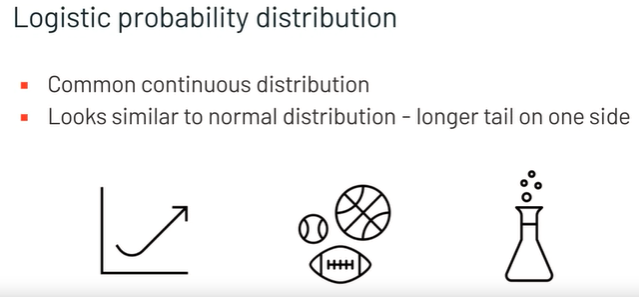
One of the most common types of continuous probability distributions is the uniform distribution. [Sürekli olasılık dağılımlarının en yaygın türlerinden biri tek biçimli dağılımdır.] This distribution describes experiments where all outcomes are equally likely, and thus their probabilities are uniformly distributed across a certain range of values between a minimum and a maximum. [Bu dağılım, tüm sonuçların eşit derecede olası olduğu deneyleri tanımlar ve bu nedenle olasılıkları, minimum ve maksimum arasındaki belirli bir değer aralığına eşit olarak dağıtılır.] Because the probabilities are the same across the entire range, the distribution looks like a rectangle and is sometimes called the rectangle distribution. [Olasılıklar tüm aralıkta aynı olduğundan, dağılım bir dikdörtgene benzer ve bazen dikdörtgen dağılımı olarak adlandırılır.] Similar to the discrete probability distribution, the probabilities in a continuous distribution all sum to one. [Kesikli olasılık dağılımına benzer şekilde, sürekli bir dağılımdaki olasılıkların toplamı bire eşittir.] That's the area under the curve between the minimum value and the maximum value equals one. [Bu, minimum değer ile maksimum değer arasındaki eğrinin altındaki alandır, bire eşittir.] An example of a uniform continuous probability distribution is a random number generator that generates random numbers between zero and one. [Tek tip sürekli olasılık dağılımına bir örnek, sıfır ile bir arasında rasgele sayılar üreten bir rasgele sayı üretecidir.]



You've probably heard of the normal distribution, often referred to as the Gaussian distribution or the bell curve. [Muhtemelen Gauss dağılımı veya çan eğrisi olarak adlandırılan normal dağılımı duymuşsunuzdur.] This is the most important probability distribution in statistics because it fits many natural phenomena. [Bu, birçok doğal olaya uyduğu için istatistikteki en önemli olasılık dağılımıdır.] The distribution is symmetric around a central peak, and the probabilities for values taper off equally on either side of that peak. [Dağılım, merkezi bir tepe etrafında simetriktir ve değerlerin olasılıkları, o tepenin her iki tarafında eşit olarak azalır.] You may recall from the beginning of this module that the standard deviation is a measure of dispersion and tells us how much are data is spread out around the mean. [Bu modülün başlangıcından, standart sapmanın bir dağılım ölçüsü olduğunu ve bize verilerin ortalamaya ne kadar dağıldığını söylediğini hatırlayabilirsiniz.] This measurement is useful and continuous distributions like the normal distribution because we can use it to describe how far a value is from the mean or the peak relative to all other values for the variable. [Bu ölçüm, normal dağılım gibi kullanışlı ve sürekli dağılımlardır, çünkü bir değerin, değişkenin diğer tüm değerlerine göre ortalamadan veya tepeden ne kadar uzakta olduğunu açıklamak için kullanabiliriz.] A larger standard deviation means that the values are more spread out, whereas a smaller standard deviation means that the values are closer to one another. [Daha büyük bir standart sapma, değerlerin daha fazla yayıldığı anlamına gelirken, daha küçük bir standart sapma, değerlerin birbirine daha yakın olduğu anlamına gelir.] An example of a normal probability distribution would be the heights of all people in a given country. [Normal bir olasılık dağılımının bir örneği, belirli bir ülkedeki tüm insanların boyları olabilir.]



There's actually a rule that relates to standard deviation to the normal distribution, the empirical rule, or the 68, 95, 99.7 rule. [Aslında normal dağılıma standart sapma, ampirik kural veya 68, 95, 99.7 kuralı ile ilgili bir kural var.] The empirical rule states that about 68% of the values in a normal distribution again, measured as the area under the curve fall within one standard deviation of the mean, about 95% of the values fall within two standard deviations of the mean, and about 99.7% of values fall within three standard deviations of the mean. [Ampirik kural, eğrinin altındaki alan olarak ölçülen normal dağılımdaki değerlerin yaklaşık %86'sının ortalamanın bir standart sapması içine düştüğünü, değerlerin yaklaşık %95'inin ortalamanın iki standart sapması içine düştüğünü ve yaklaşık olarak yaklaşık %86'sının olduğunu belirtir. Değerlerin %99,7'si ortalamanın üç standart sapması içinde yer alır.] With a normal distribution, most outcomes will be within three standard deviations of the mean. [Normal bir dağılımla, çoğu sonuç, ortalamanın üç standart sapması içinde olacaktır.]



The logistic probability distribution is another common continuous distribution. [Lojistik olasılık dağılımı, başka bir yaygın sürekli dağılımdır.] The shape of the distribution is defined by the logistic function, which will bring up later in the course. [Dağıtımın şekli, kursta daha sonra ortaya çıkacak olan lojistik fonksiyon tarafından tanımlanır.] It looks similar to the normal distribution and that its symmetric and has one peak, but it has a longer tail on one side. [Normal dağılıma benziyor ve simetrik ve bir tepesi var, ancak bir tarafında daha uzun bir kuyruğu var.] It's frequently used to model growth and in sports, physical sciences and finance. [Büyümeyi modellemek için ve spor, fizik bilimleri ve finansta sıklıkla kullanılır.] Is a better fit than the normal distribution for extreme events or outliers. [Aşırı olaylar veya aykırı değerler için normal dağılımdan daha uygundur.] An example of a logistic probability distribution is the salaries of professional athletes. [Lojistik olasılık dağılımına bir örnek, profesyonel sporcuların maaşlarıdır.] You should have a good understanding of how probability distributions are used to model data. [Verileri modellemek için olasılık dağılımlarının nasıl kullanıldığını iyi anlamalısınız.] The difference between discrete and continuous random variables and some of the most common probability distributions and their applications. [Kesikli ve sürekli rastgele değişkenler arasındaki fark ve en yaygın olasılık dağılımlarından bazıları ve uygulamaları.]