**Introductory Statistics for Data Science**

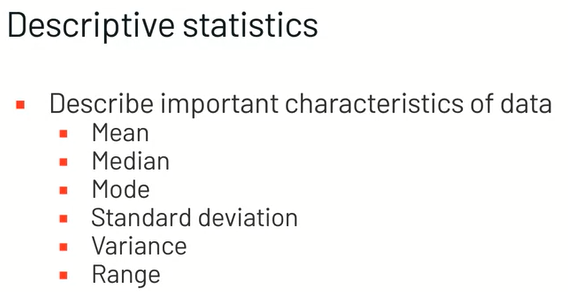
**Learning Objectives**

* 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]

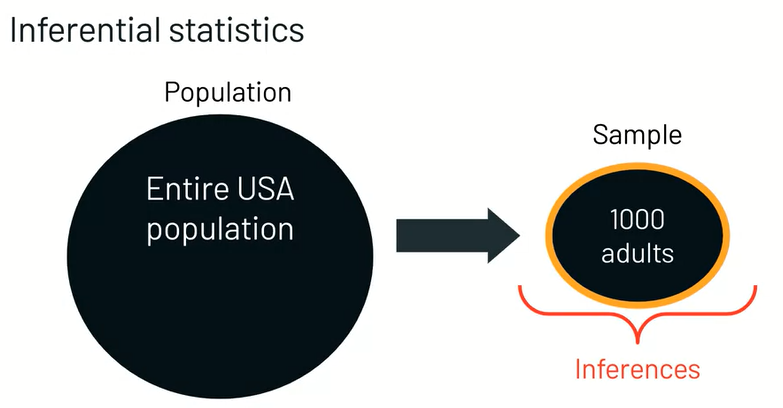
# **Probability Distributions**

# **Statistics and Probability Review**

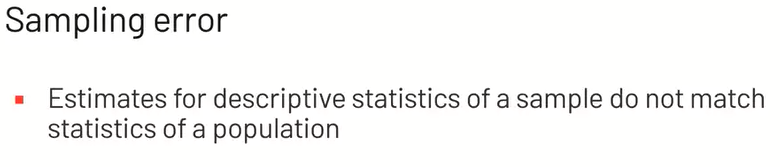
Hello, and welcome back. [Merhaba ve tekrar hoş geldiniz.] In this lesson, we will learn about probability distributions and how they can be used to model data. [Bu derste, olasılık dağılımlarını ve bunların verileri modellemek için nasıl kullanılabileceğini öğreneceğiz.] But first, let's recap what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



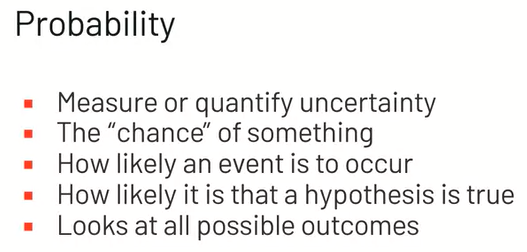
Remember that we can take a sample, or subset of data from some population, and calculate descriptive statistics to describe that sample. [Bir popülasyondan bir örnek veya veri alt kümesi alabileceğimizi ve bu örneği tanımlamak için tanımlayıcı istatistikleri hesaplayabileceğimizi unutmayın.] These include 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.]

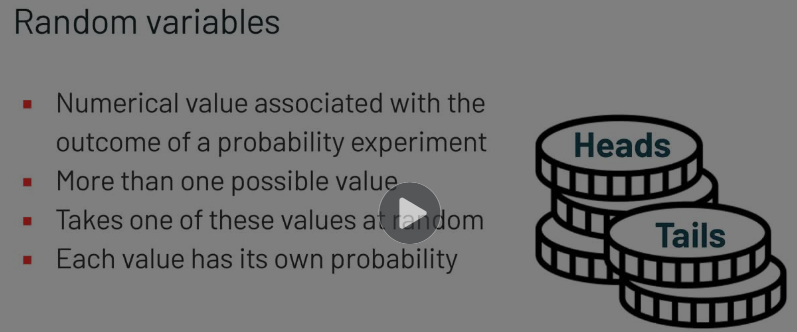


Even so, we know that there will be some sampling error, meaning that our estimates for the descriptive statistics of the sample won't perfectly match the statistics of the population. [Öyle olsa bile, bir miktar örnekleme hatası olacağını biliyoruz; bu, örneğin tanımlayıcı istatistiklerine ilişkin tahminlerimizin, 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.]