**Introductory Statistics for Data Science**

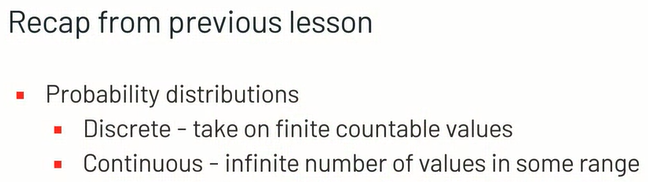
**Learning Objectives**

* 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]

# **Applying Statistics to Learn From Data**

# **Probability Distribution Review**

In this lesson, will continue building on top of what we've already learned about statistical methods, and probability distributions. [Bu derste, istatistiksel yöntemler ve olasılık dağılımları hakkında öğrendiklerimizin üzerine inşa etmeye devam edeceğiz.] And will begin to explore how we can use these techniques to learn from data and generalize to the real world. [Ve verilerden öğrenmek ve gerçek dünyaya genellemek için bu teknikleri nasıl kullanabileceğimizi keşfetmeye başlayacak.] Before we move on to some new concepts, let's recap what we covered in the previous lesson about probability distributions, random variables, and how to use them to model data. [Bazı yeni kavramlara geçmeden önce, olasılık dağılımları, rasgele değişkenler ve bunların verileri modellemek için nasıl kullanılacağı hakkında önceki derste ele aldıklarımızı özetleyelim.]



Probability distributions are used to describe and display the probabilities associated with the values of a random variable that are the outcomes of some type of experiment. [Olasılık dağılımları, bir tür deneyin sonucu olan rastgele bir değişkenin değerleriyle ilişkili olasılıkları tanımlamak ve görüntülemek için kullanılır.] These random variables can be discrete or they take on finite countable values or continuous where they can have an infinite number of values in some range. [Bu rastgele değişkenler ayrık olabilir veya sonlu sayılabilir değerler alabilir veya bir aralıkta sonsuz sayıda değere sahip olabilecekleri sürekli olabilir.] A probability distribution maps these values to their probabilities. [Bir olasılık dağılımı, bu değerleri olasılıklarına göre eşleştirir.] They have certain properties that describe different qualities of the data set and are important for making inferences. [Veri setinin farklı niteliklerini tanımlayan belirli özelliklere sahiptirler ve çıkarım yapmak için önemlidirler.] Some of the discrete probability distributions we learned about in the last lesson are Binomial, Multinomial, Poisson and Bernoulli. [Son derste öğrendiğimiz kesikli olasılık dağılımlarından bazıları Binom, Multinomial, Poisson ve Bernoulli'dir.] Some of the continuous distributions we talked about our uniform, normal and logistic. [Üniforma, normal ve lojistik hakkında konuştuğumuz sürekli dağılımlardan bazıları.] Okay, now we'll move on and see how all this pieces are connected. [Tamam, şimdi devam edeceğiz ve tüm bu parçaların nasıl bağlantılı olduğunu göreceğiz.]

# **An Introduction to Hypothesis Testing**

You may be wondering how all the things you've learned about statistics, probability, and probability distributions fit into data science. [İstatistikler, olasılıklar ve olasılık dağılımları hakkında öğrendiğiniz her şeyin veri bilimine nasıl uyduğunu merak ediyor olabilirsiniz.] These tools are actually foundational to how data science works and how we can use data to make inferences and predictions about the real world. [Bu araçlar aslında veri biliminin nasıl çalıştığının ve gerçek dünya hakkında çıkarımlar ve tahminler yapmak için verileri nasıl kullanabileceğimizin temelidir.] Previously in this course, we talked about the scientific method. [Bu derste daha önce bilimsel yöntemden bahsetmiştik.] One of the steps involved formulating a hypothesis and then assessing it, usually by either rejecting or failing to reject the null hypothesis. [Adımlardan biri, bir hipotezin formüle edilmesini ve ardından, genellikle sıfır hipotezini reddederek veya reddetmeyerek onu değerlendirmeyi içeriyordu.] Now in this module, we'll actually learn how statistical tools are used to do this. [Şimdi bu modülde, aslında bunu yapmak için istatistiksel araçların nasıl kullanıldığını öğreneceğiz.] First, let's dive in a little deeper on what exactly a hypothesis is. [İlk olarak, tam olarak bir hipotezin ne olduğu konusunda biraz daha derine inelim.] A hypothesis is a way to formulate a question or an idea about something in a way that is statistically testable through data collection, analysis and inference. [Bir hipotez, veri toplama, analiz ve çıkarım yoluyla istatistiksel olarak test edilebilir bir şekilde bir soru veya bir şey hakkında bir fikir formüle etmenin bir yoludur.] We already know that an experimental hypothesis is formulated in two parts, the null hypothesis and the alternative hypothesis. [Deneysel bir hipotezin, sıfır hipotezi ve alternatif hipotez olmak üzere iki kısımda formüle edildiğini zaten biliyoruz.] The null hypothesis is usually associated with an expectation that something will have no impact or no effect, such as a new medication having no impact on life expectancy. [Sıfır hipotezi, genellikle, yeni bir ilacın yaşam beklentisi üzerinde hiçbir etkisi olmaması gibi, bir şeyin etkisinin olmayacağı veya hiçbir etkisinin olmayacağı beklentisiyle ilişkilidir.] The alternative hypothesis is usually associated with an expectation that something will happen, impact or effect, such as a new medication, increasing life expectancy. [Alternatif hipotez genellikle bir şeyin olacağı, etki veya etki, örneğin yeni bir ilaç, artan yaşam beklentisi gibi bir beklenti ile ilişkilidir.] We can either reject or fail to reject the null hypothesis, but we don't ever say that we accept the alternative hypothesis. [Sıfır hipotezini reddedebilir veya reddedebiliriz, ancak alternatif hipotezi kabul ettiğimizi asla söylemiyoruz.] This is because we can't ever say for certain that our hypothesis is true for the entire population, because we can't study it to verify your claim. [Bunun nedeni, hipotezimizin tüm popülasyon için doğru olduğunu kesin olarak söyleyemememizdir, çünkü iddianızı doğrulamak için onu inceleyemeyiz.] Formulating a hypothesis with an alternative and a null allows us to evaluate it using statistical tools, specifically hypothesis testing. [Bir alternatif ve bir boş değerle bir hipotez formüle etmek, onu istatistiksel araçlar, özellikle de hipotez testi kullanarak değerlendirmemize izin verir.] The reason we need hypothesis testing is because the patterns we observe in a sample could be due to sampling error rather than being representative of a true effect in the population. [Hipotez testine ihtiyaç duymamızın nedeni, bir örneklemde gözlemlediğimiz kalıpların, popülasyondaki gerçek bir etkiyi temsil etmekten ziyade örnekleme hatasından kaynaklanabilmesidir.] If this is the case, the observed patterns might be different the next time the experiment is performed or the sample is collected. [Bu durumda, gözlemlenen modeller, deneyin bir sonraki gerçekleştirilmesinde veya numune toplandığında farklı olabilir.] Hypothesis testing uses estimates of the sampling error to make an accurate inference from the sample to the population. [Hipotez testi, örneklemden popülasyona doğru bir çıkarım yapmak için örnekleme hatası tahminlerini kullanır.] The main function of statistics is actually to test hypothesis. [İstatistiğin temel işlevi aslında hipotezi test etmektir.] Generally speaking, we do this by comparing the null and alternative hypothesis, and then using a defined formula to determine if the difference between their probability distributions is statistically significant based on whether the difference falls below a predetermined threshold value which is called the significance level. [Genel olarak konuşursak, bunu boş ve alternatif hipotezi karşılaştırarak ve ardından olasılık dağılımları arasındaki farkın istatistiksel olarak anlamlı olup olmadığını belirlemek için tanımlanmış bir formül kullanarak, farkın anlamlılık düzeyi olarak adlandırılan önceden belirlenmiş bir eşik değerinin altına düşüp düşmediğine bağlı olarak yaparız.] In most cases, we refer to this as the P-value. [Çoğu durumda, buna P değeri diyoruz.] If it does fall below this significance level, we can say that the patterns we have observed in our sample data are very unlikely to occur if the null hypothesis were true. [Bu anlamlılık seviyesinin altına düşerse, sıfır hipotezi doğru olsaydı, örnek verilerimizde gözlemlediğimiz kalıpların ortaya çıkma ihtimalinin çok düşük olduğunu söyleyebiliriz.] Therefore, we reject the null hypothesis and can say that the results are statistically significant. [Bu nedenle, boş hipotezi reddediyoruz ve sonuçların istatistiksel olarak anlamlı olduğunu söyleyebiliriz.] In short, hypothesis testing as a procedure that uses statistical techniques to test an assumption about a population based on sample data. [Kısacası, hipotez testi, örnek verilere dayalı bir popülasyon hakkında bir varsayımı test etmek için istatistiksel teknikleri kullanan bir prosedür olarak.] Let's look at the four steps involved in conducting a hypothesis test. [Bir hipotez testi yürütmekle ilgili dört adıma bakalım.] Step one, we state the null and alternative hypothesis, they must be mutually exclusive, meaning only one of them can be true. [Birinci adım, sıfır ve alternatif hipotezi belirtiyoruz, bunlar birbirini dışlamalı, yani bunlardan sadece biri doğru olabilir.] Step two, we identify a test statistic that will use to evaluate the null hypothesis. [İkinci adım, sıfır hipotezini değerlendirmek için kullanılacak bir test istatistiği tanımlarız.] In step three, we analyzed the sample data and find the value of the test statistic. [Üçüncü adımda, örnek verileri analiz ettik ve test istatistiğinin değerini bulduk.] In step four, we interpret the results an reject or fail to reject the null hypothesis. [Dördüncü adımda, sonuçları sıfır hipotezini reddet veya reddet olarak yorumluyoruz.] This is usually done by comparing the calculated P-value with an acceptable significance value. [Bu genellikle hesaplanan P-değerinin kabul edilebilir bir anlamlılık değeriyle karşılaştırılmasıyla yapılır.] This is most often called alpha and we set it at 0.05 by convention. [Bu genellikle alfa olarak adlandırılır ve biz geleneksel olarak 0,05 olarak belirledik.] That was a lot of information, but don't worry, this was just meant to give you an overview of some of the ideas and terminology that will be using in this lesson. [Bu çok fazla bilgiydi, ama merak etmeyin, bu sadece size bu derste kullanılacak bazı fikirler ve terminoloji hakkında genel bir bilgi vermek içindi.] The main idea to take away is that hypothesis testing is a way to determine whether the results of a survey or experiment are meaningful. [Çıkarılması gereken ana fikir, hipotez testinin bir anketin veya deneyin sonuçlarının anlamlı olup olmadığını belirlemenin bir yolu olmasıdır.] In the next video, we'll go through an example of how to use hypothesis testing to answer a question. [Bir sonraki videoda, bir soruyu yanıtlamak için hipotez testinin nasıl kullanılacağına dair bir örnek üzerinden gideceğiz.]