Unlike the other hypothesis testing of science, statistical hypothesis testing always involves assessing an assertion made about the value of a population parameter. This assertion is stated as a pair of statements: the null hypothesis (Ho) and the alternative hypothesis (Ha or H1).

Bilimin diğer hipotez testlerinden farklı olarak, istatistiksel hipotez testi her zaman bir popülasyon parametresinin değeri hakkında yapılan bir iddianın değerlendirilmesini içerir. Bu iddia bir çift ifade olarak ifade edilir: boş hipotez (Ho) ve alternatif hipotez (Ha veya H1).



The **null hypothesis** always expresses equality and it is the thing we are trying to provide evidence against. However, the **alternative hypothesis** is usually the idea we are studying about and typically it is the thing that we want to prove.

Sıfır hipotezi her zaman eşitliği ifade eder ve bizim aleyhine kanıt sağlamaya çalıştığımız şeydir. Bununla birlikte, alternatif hipotez genellikle üzerinde çalıştığımız fikirdir ve genellikle kanıtlamak istediğimiz şeydir.

**Hypothesis**  
In statistics, a hypothesis is a statement about a population, usually claiming that a parameter takes a particular numerical value or falls in a certain range of values.

İstatistikte hipotez, bir popülasyon hakkında, genellikle bir parametrenin belirli bir sayısal değer aldığını veya belirli bir değer aralığına düştüğünü iddia eden bir ifadedir.

Before conducting a hypothesis testing, we identify the variable measured and the population parameter of interest. For a categorical variable the parameter is a proportion, and for a quantitative variable the parameter is a mean.

Bir hipotez testi yapmadan önce, ölçülen değişkeni ve ilgilenilen popülasyon parametresini tanımlarız. Kategorik bir değişken için parametre bir orantıdır ve nicel bir değişken için parametre bir ortalamadır.

Basic Concepts of Hypothesis Testing

The Steps of a Hypothesis Testing

A significance test has five steps. In this section, we introduce the general ideas behind these steps.

Bir anlamlılık testinin beş adımı vardır. Bu bölümde, bu adımların arkasındaki genel fikirleri tanıtıyoruz.

**1. Assumptions**

First, specify the variable and parameter. The assumptions commonly pertain to the method of data production (randomization), the sample size, and the shape of the population distribution.

İlk olarak, değişkeni ve parametreyi belirtin. Varsayımlar genellikle veri üretim yöntemi (rastgeleleştirme), örneklem büyüklüğü ve popülasyon dağılımının şekli ile ilgilidir.

### 2. Hypotheses

State the null hypothesis, H0 (a single parameter value, usually no effect), and the alternative hypothesis, Ha (a set of alternative parameter values)

2. Hipotezler

Boş hipotezi, H0'ı (tek bir parametre değeri, genellikle etkisi yoktur) ve alternatif hipotezi, Ha'yı (bir dizi alternatif parametre değeri) belirtin.

Example-1

Null Hypotheses :

The population mean time to answer customer complaints was 5 minutes in 2020.

Örnek 1

Boş Hipotezler:

Popülasyonun müşteri şikayetlerini cevaplama süresi mean’i 2020'de 5 dakika idi.

Alternative Hypotheses :

The population mean time to answer customer complaints was not 5 minutes in 2020.

Alternatif Hipotezler:

Popülasyonun müşteri şikayetlerini cevaplama süresi mean’i 2020'de 5 dakika değildi.

Example-2

Null Hypotheses :

The average weight for men is the same as the average weight for women in USA.

Boş Hipotezler:

Erkeklerin ortalama ağırlığı, ABD'deki kadınların ortalama ağırlığı ile aynıdır.

Alternative Hypotheses :

The average weight for men is not the same as the average weight for women in USA.

Alternatif Hipotezler:

Erkekler için ortalama ağırlık, ABD'deki kadınların ortalama ağırlığı ile aynı değildir.

### 3. Test Statistic

The parameter to which the hypotheses refer has a point estimate. A test statistic describes how far that point estimate falls from the parameter value given in the null hypothesis. Usually this distance is measured by the number of standard errors between the point estimate and the parameter.

Hipotezlerin atıfta bulunduğu parametrenin bir nokta tahmini vardır. Bir test istatistiği, bu nokta tahmininin boş hipotezde verilen parametre değerinden ne kadar uzak olduğunu açıklar. Genellikle bu mesafe, nokta tahmini ile parametre arasındaki standart hataların sayısı ile ölçülür.

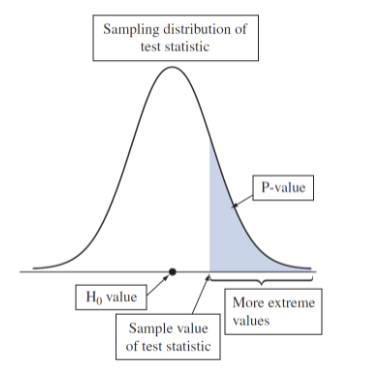


HOCA: test istatistiklerinde; popülasyon standart sapmalarından (sigma) oluşturacağımız standart hatayı kullanıyoruz.

### 4. P-value

The P-value is the probability that the test statistic takes the observed value or a value more extreme if we presume H0 is true. Smaller P-values represent stronger evidence against H0.

P-değeri, test istatistiğinin gözlemlenen değeri veya H0'ın doğru olduğunu varsayarsak daha uç bir değeri alma olasılığıdır. Daha küçük P değerleri, H0'a karşı daha güçlü kanıtları temsil eder.



Test istatistiği verisinden yola çıkarak bir P değeri hesabı yapıyoruz. Yani bu test istatistiğinden hareketle bir extreme değerlere ulaşıyoruz. Bu extreme values’e P-values deniyor.

Düşük p değeri H0’ı reddetmekte daha güçlü kanıt anlamına geliyor. Çünkü 0 hipotezinde belirttiğiniz parametre değerinden (mesela mean) daha fazla uzaklaştığın anlamına geliyor. Test istatistiğin H0 parametresinden ne kadar uzaklaşırsa o kadar güçlü kanıt buldun demektir.

**5. Conclusion**

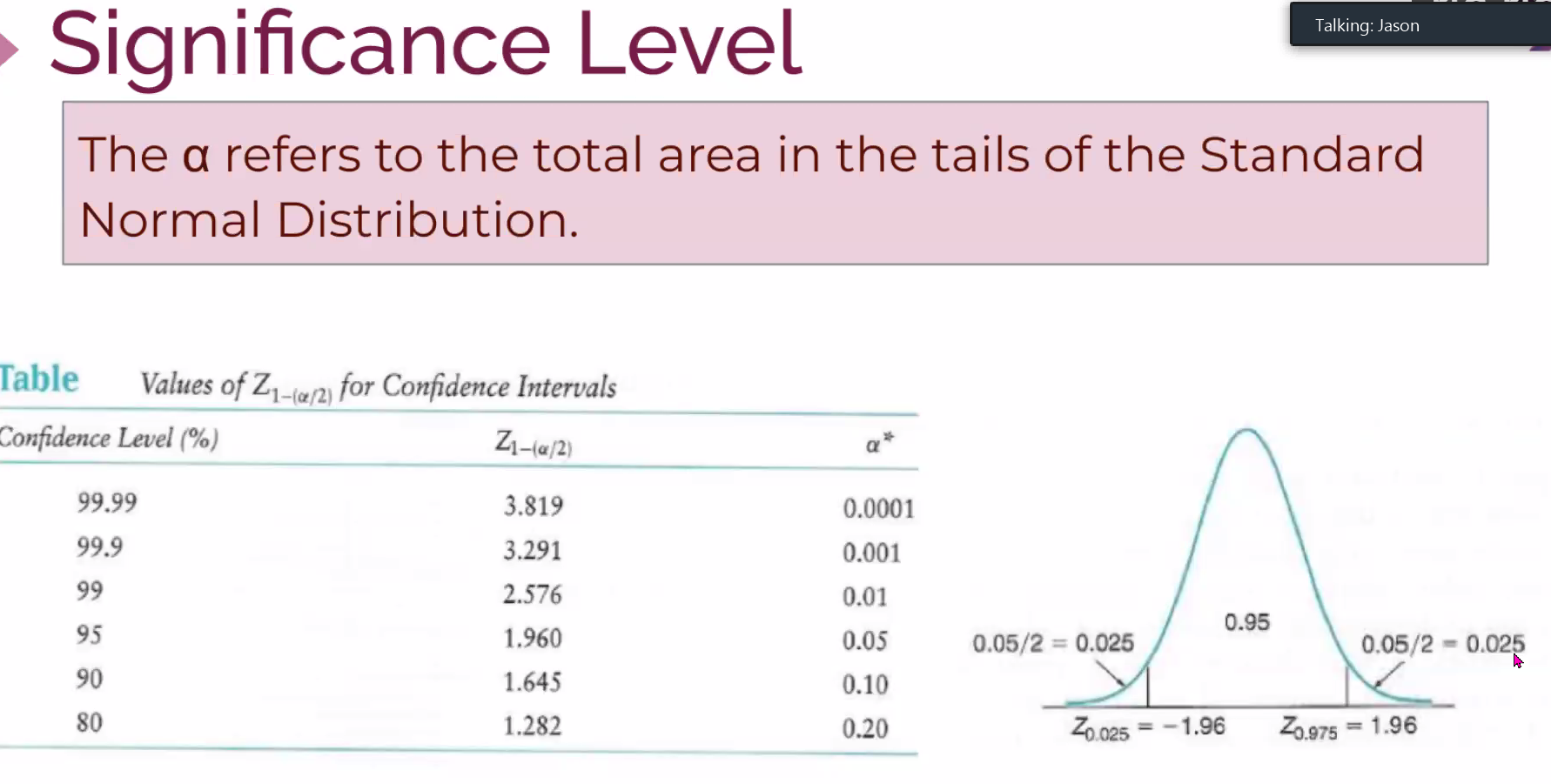
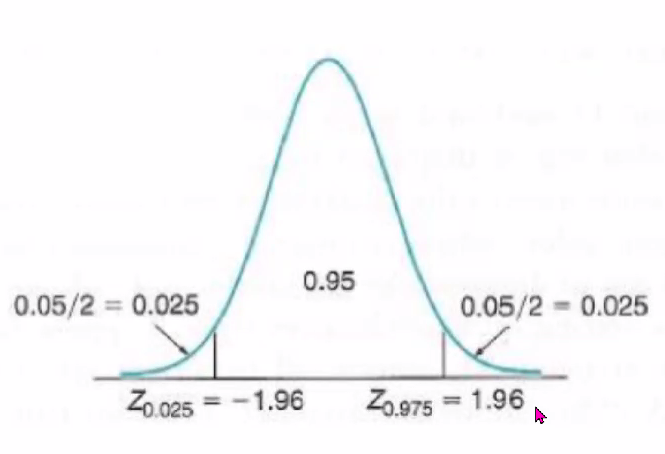
Report and interpret the P-value in the context of the study. Based on the P-value, make a decision about H0 (either reject or do not reject H0 ) if a decision is needed.

P değeriniz büyürse

İşin içine P haricinde significance level (anlamlılık derecesi) giriyor. Ortalama değerimiz var ve bir test istatistiği uyguluyoruz (mesela z). Soru şu: P value ne kadar küçük olmalı ki reject için yeterli bir kanıt oldu diyebileyim? Öyle bir z değeri bulacağım ki test istatistiği ile p değerini karşılaştıracağım ve p’yi yeterince düşük bulacağım. Sonuçta bu sayede 0 hipotezini reddedeceğim veya reddedemeyeceğim. (Reject H0 if P-value <= 0.05)

95% Cİ aslında 0.05 anlamlılık seviyesine karşılık gelir.

1. Ci = 0.05

**P-value’yu bulmak için test istatistiğinden yola çıkıyoruz. P-value bir istatistiktir.**

5. Sonuç

P-değerini çalışma bağlamında raporlayın ve yorumlayın. P değerine dayanarak, bir karar gerekiyorsa H0 hakkında bir karar verin ( H0 reddet ya da reddet).



## Basic Concepts of Hypothesis Testing

### One and Two Tailed Tests

**Two-tailed:** Sometimes the Rejection Region is broken into two parts and lies in both tails of the sampling distribution of the test statistic (when H0:μ = μ0 is true)

İki kuyruklu: Bazen Reddetme Bölgesi iki parçaya bölünür ve test istatistiğinin örnekleme dağılımının her iki kuyruğunda bulunur (H0:μ = μ0 doğru olduğunda)

H0: "Drug has no effect (μ = μ0)"

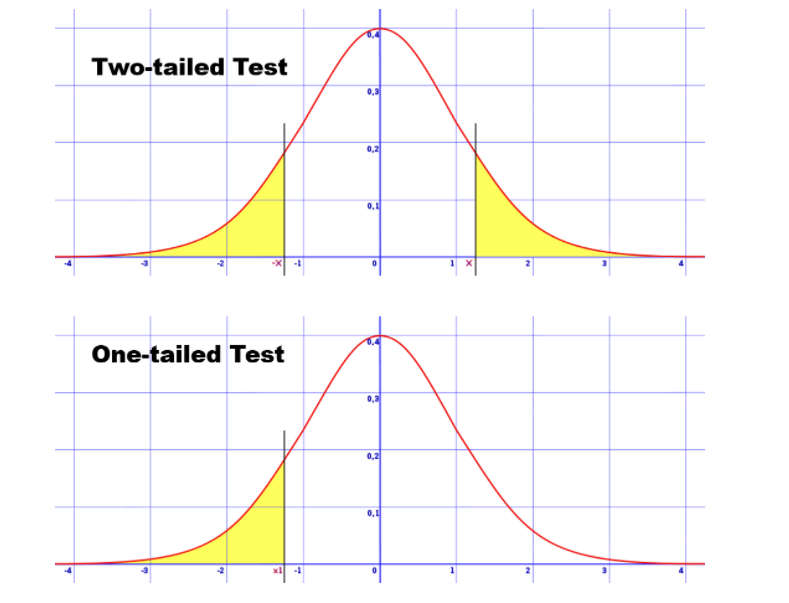
H1: "Drug has an effect (μ ≠ μ0)"

**One-tailed:** Sometimes the Rejection Region lies entirely in one tail of the sampling distribution of the test statistic (when H0:μ = μ0 is true versus H1:μ > μ0).

Tek kuyruklu: Bazen Reddetme Bölgesi, test istatistiğinin örnekleme dağılımının bir kuyruğunda yer alır (H0:μ = μ0, H1:μ > μ0'a karşı doğru olduğunda).

H0: "Drug has no effect (μ = μ0)"

H1: "Drug lowers response time (μ < μ0)"



## Basic Concepts of Hypothesis Testing

### Significance Level and P-Value

### Significance Level

The significance level (alpha or α), is a measure of the strength of the evidence to be present in the sample before you will reject the null hypothesis and conclude that the effect is statistically significant. Before conducting the experiment, the researcher determines the significance level. And we can say the typical values for the significance level are 0.1, 0.05, and 0.01.

Önem Düzeyi

Anlamlılık düzeyi (alfa veya α), boş hipotezi reddetmeden ve etkinin istatistiksel olarak anlamlı olduğu sonucuna varmadan önce örnekte bulunacak kanıtın gücünün bir ölçüsüdür. Araştırmacı deneyi gerçekleştirmeden önce anlamlılık düzeyini belirler. Ve anlamlılık düzeyi için tipik değerlerin 0.1, 0.05 ve 0.01 olduğunu söyleyebiliriz.

In other words, **the significance level is the probability of rejecting the null hypothesis when it is true.** For instance, a significance level of 0.05 represents a 5% risk of concluding that a difference exists when there is no actual difference. A lower significance level indicates that we require stronger evidence before rejecting the null hypothesis.

Başka bir deyişle, anlamlılık düzeyi, sıfır hipotezinin doğru olduğunda reddedilme olasılığıdır. Örneğin, 0,05 anlamlılık düzeyi, gerçek bir fark olmadığında bir farkın var olduğu sonucuna varma riskinin %5'ini temsil eder. Daha düşük bir anlamlılık düzeyi, sıfır hipotezini reddetmeden önce daha güçlü kanıtlara ihtiyacımız olduğunu gösterir.

### P-Value

A p-value is the probability that we would obtain the effect observed in our sample, if the null hypothesis is true for the population. P-value is calculated based on our sample data and under the assumption that the null hypothesis is true. A lower p-value indicates greater evidence against the null hypothesis.

Bir p değeri, popülasyon için sıfır hipotezi doğruysa, örneğimizde gözlemlenen etkiyi elde etme olasılığımızdır. P-değeri, örnek verilerimize dayanarak ve boş hipotezin doğru olduğu varsayımı altında hesaplanır. Daha düşük bir p değeri, boş hipoteze karşı daha fazla kanıt olduğunu gösterir.

Using the significance level and p-value during hypothesis testing helps us determine which hypothesis the data support. We compare the p-value with the significance level. If the p-value is less than our significance level, we can reject the null hypothesis and conclude that the effect is statistically significant. In other words, the evidence in our sample is strong enough to be able to reject the null hypothesis at the population level.

Hipotez testi sırasında anlamlılık düzeyini ve p-değerini kullanmak, verilerin hangi hipotezi desteklediğini belirlememize yardımcı olur. p-değerini anlamlılık düzeyi ile karşılaştırırız. p değeri bizim anlamlılık seviyemizden küçükse, sıfır hipotezini reddedebilir ve etkinin istatistiksel olarak anlamlı olduğu sonucuna varabiliriz. Başka bir deyişle, örneğimizdeki kanıtlar popülasyon düzeyinde sıfır hipotezini reddedebilecek kadar güçlüdür.

## Basic Concepts of Hypothesis Testing

### Type-I and Type-II Errors

A hypothesis test is not 100% reliable because it uses a random sample to draw a conclusion about the entire population. When we perform a hypothesis test, there are **two types of errors** related to drawing an incorrect conclusion.

**Type I error:** That rejects a null hypothesis although it is true. We can think of this as a false positive.  
**Type II error:** That test fails to reject a null hypothesis although it is false. We can think of this as a false negative

Hipotez Testinin Temel Kavramları

Tip-I ve Tip-II Hatalar

Bir hipotez testi, tüm popülasyon hakkında bir sonuç çıkarmak için rastgele bir örnek kullandığından %100 güvenilir değildir. Bir hipotez testi yaptığımızda, yanlış bir sonuca varmakla ilgili iki tür hata vardır.

Tip I hata: Bu, doğru olmasına rağmen boş bir hipotezi reddeder. Bunu yanlış bir pozitif olarak düşünebiliriz.

O hipotezinin doğru olduğunu biliyorum (kadın hamile) ama yanlış test sonuçlarıyla bunun yanlış olduğunu ispatlıyorum (ispatladığımı düşünüyorum.)

ve hamile değil sonucu çıkıyor

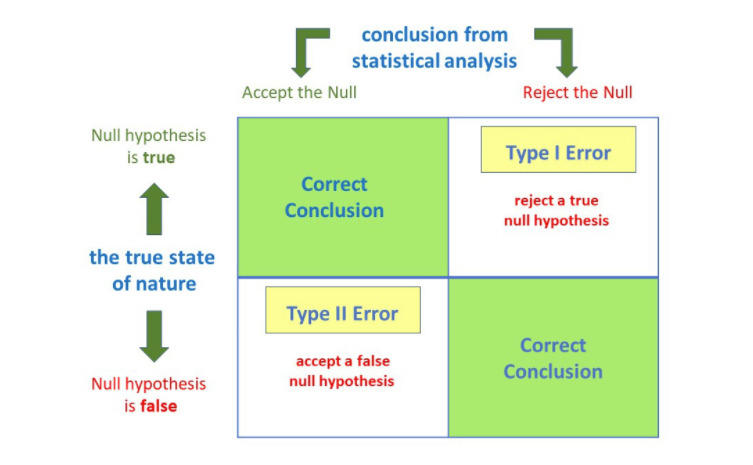
Tip II hata: Test, yanlış olmasına rağmen boş bir hipotezi reddetmede başarısız olur. Bunu yanlış bir negatif olarak düşünebiliriz.

0 hipotezinin (hamile değil hipotezi) yanlış olduğunu biliyoruz ama yanlış test sonuçlarıyla 0 hipotezini reddedemiyoruz (hamile olve yanlış olanı doğru kabul ediyoruz.

Tip 1 Hata: Doğru olan 0 hipotezini reddetmemeniz gerekirken, elindeki test verileri ile reddediyorsun.

Test 2 Hata: Yanlış olan 0 hipotezini reddetmeniz gerekirken test değerleri ile reddedemiyorsun

Type 2 = suçlunun salıverilmesi, type1= masum birinin cezalandırılması.. (akıl yanılabiliyor..)



## Hypothesis Tests

### Independent Samples T Test

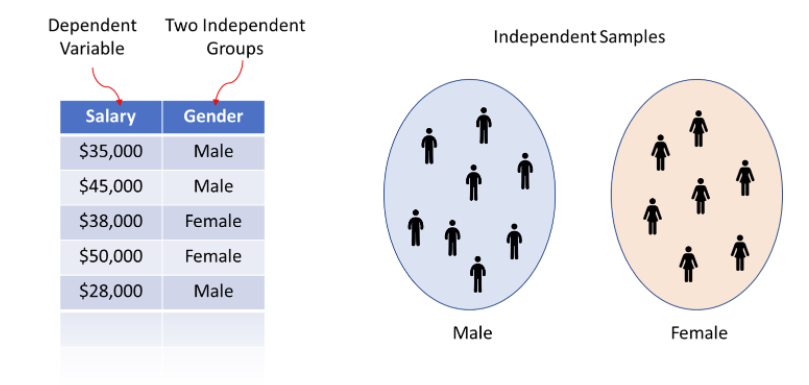
The independent-samples t-test (or independent t-test) compares the means between two independent groups on the same continuous, dependent variable.

For example, we can answer the following research question by using the independent-samples t-test.

Bağımsız örnekler t testi (veya bağımsız t testi), aynı sürekli, bağımlı değişken üzerinde iki bağımsız grup arasındaki ortalamaları karşılaştırır.

Örneğin aşağıdaki araştırma sorusunu bağımsız örnekler t-testini kullanarak cevaplayabiliriz.

Does first year graduate salaries differ based on gender? İlk yıl mezun maaşları cinsiyete göre farklılık gösteriyor mu?



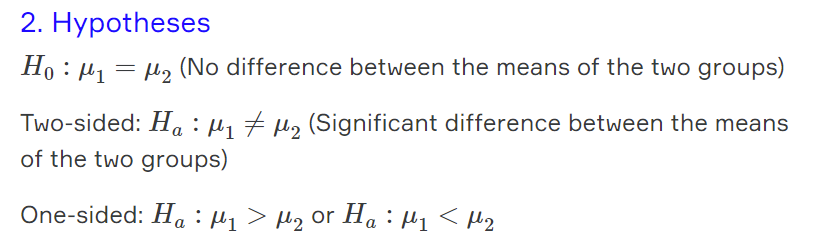
Dependent variable would be "first year graduate salaries" and independent variable would be "gender", which has two groups: "male" and "female"

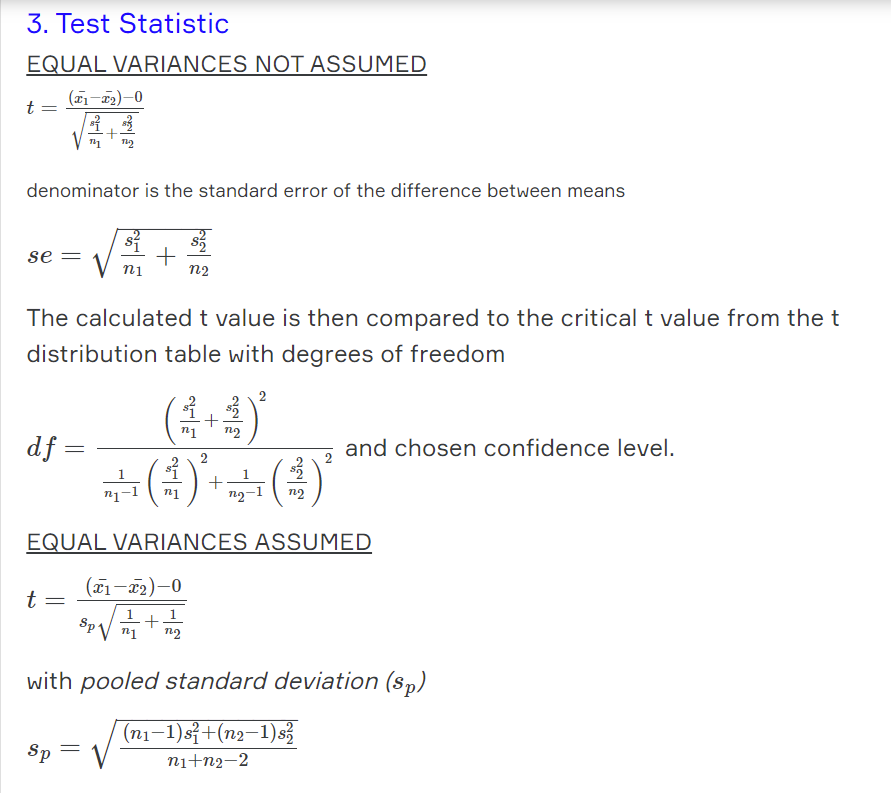
Bağımlı değişken "birinci yıl mezun maaşları" ve bağımsız değişken "erkek" ve "kadın" olmak üzere iki gruptan oluşan "cinsiyet" olacaktır.



#### 1. Assumptions

* A quantitative response variable for two groups
* Independent random samples, either from random sampling or a randomized experiment
* Approximately normal population distribution for each group. (This is mainly important for small sample sizes, and even then the two-sided test is robust to violations of this assumption.)
* 1. Varsayımlar
* İki grup için nicel bir yanıt değişkeni
* Rastgele örneklemeden veya rastgele bir deneyden bağımsız rastgele örnekler
* Her grup için yaklaşık olarak normal nüfus dağılımı. (Bu esas olarak küçük örneklem boyutları için önemlidir ve o zaman bile iki taraflı test bu varsayımın ihlaline karşı sağlamdır.)





The calculated *t* value is then compared to the critical; *t* value from the *t* distribution table with degrees of freedom; *df* = *n*1 + *n*2 - 2 and chosen confidence level.

Hesaplanan t değeri daha sonra kritik değerle karşılaştırılır; serbestlik dereceli t dağılım tablosundan t değeri; df = n1 + n2 - 2 ve seçilen güven düzeyi.

#### 4. P-value

P-value = Two-tail probability from t distribution of values even more extreme than observed t test statistic, presuming the null hypothesis is true with df given by software.

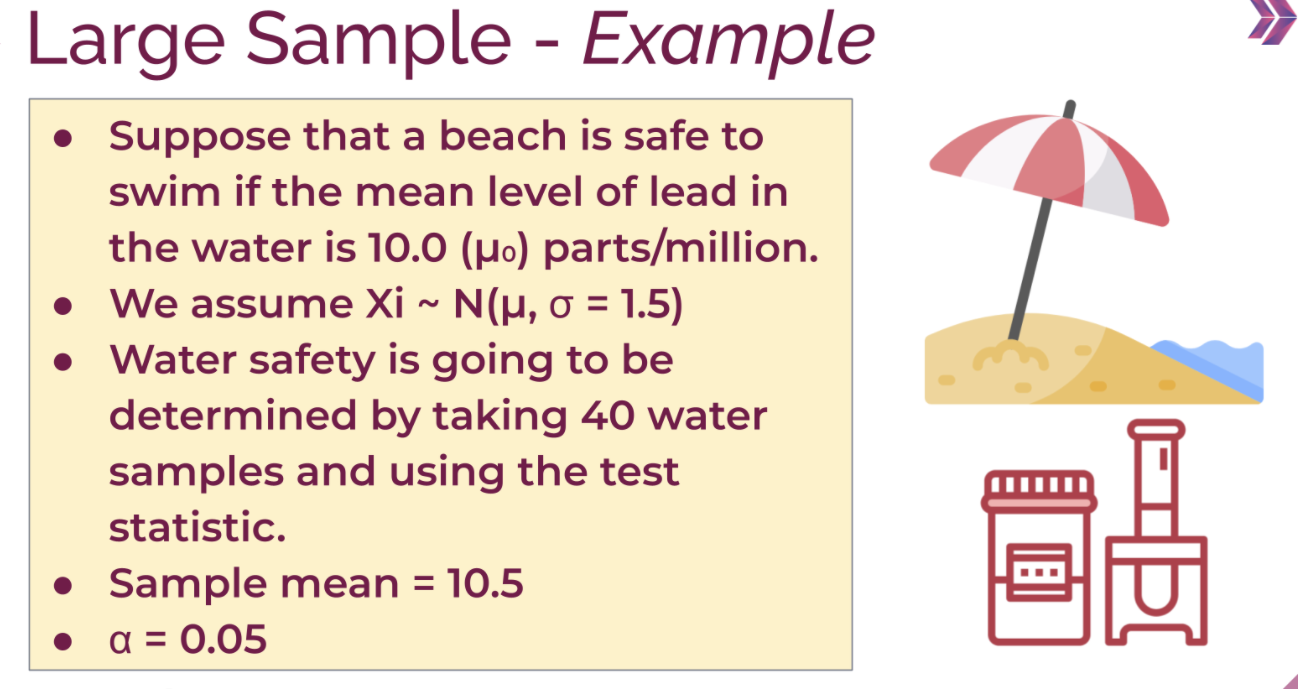
P-değeri = Yazılım tarafından verilen df ile boş hipotezin doğru olduğunu varsayarak, gözlemlenen t testi istatistiğinden bile daha aşırı değerlerin t dağılımından iki kuyruklu olasılık.

#### 5. Conclusion

Smaller P-values give stronger evidence against H0H0 and supporting HaHa. Interpret the P-value in context, and if a decision is needed, reject H0H0 if P-value ≤ significance level (such as 0.05).

**💡Tips:**

* ***Levene’s test*** is commonly used to check for homogeneity of variance.
* If Levene’s test indicates that the variances are equal across the two groups (i.e., p-value large), we will assume Equal variances.



İnsanların yüzdüğü bir sahilde kurşun seviyesi ölçümü yapılıyor. Milyon başına 10 parça kurşun benim bilinen 0 hipotezim.

Elimizdeki sample veri setine ilişkin std biliyoruz mean’i biliyoruz.

Sahilin farklı farlı noktlarından 40 tane numune alınıyor ve test istatistiği uyguluyor

Significant (alfa) 0.05 olduğu söyleniyor. (Söylemeseydi de default olarak 0.05 alınacaktı)

Assumption

random seçildi mi evet

Gözlemlerim bağımsız mı

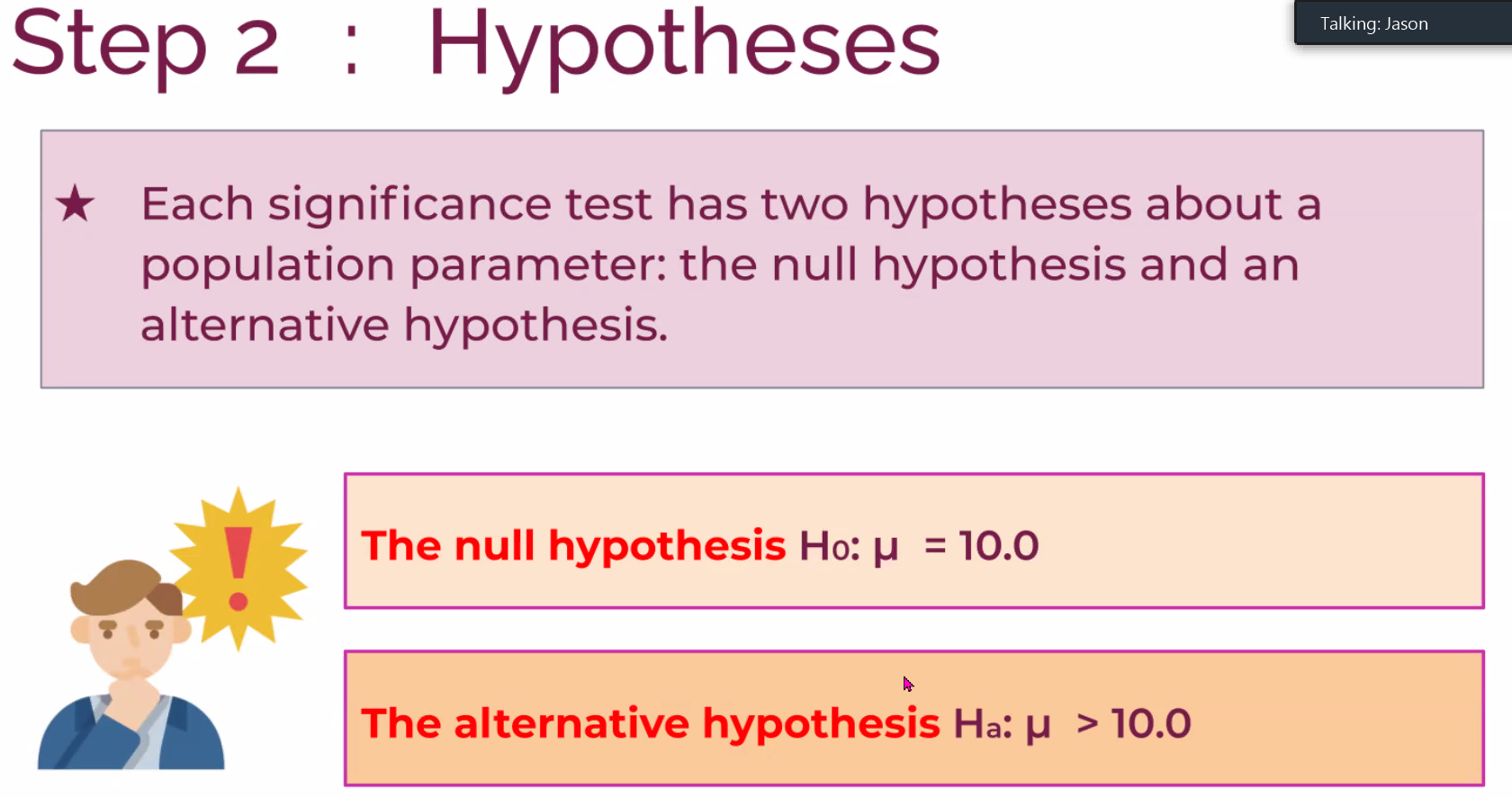
Std biliyor myum evet

Örneklemim 30 dan fazla mı fazla

Demek ki standart dağılımı (z dağılımı) kullanacağım

0 hipotezim 🡪 Mü = 10

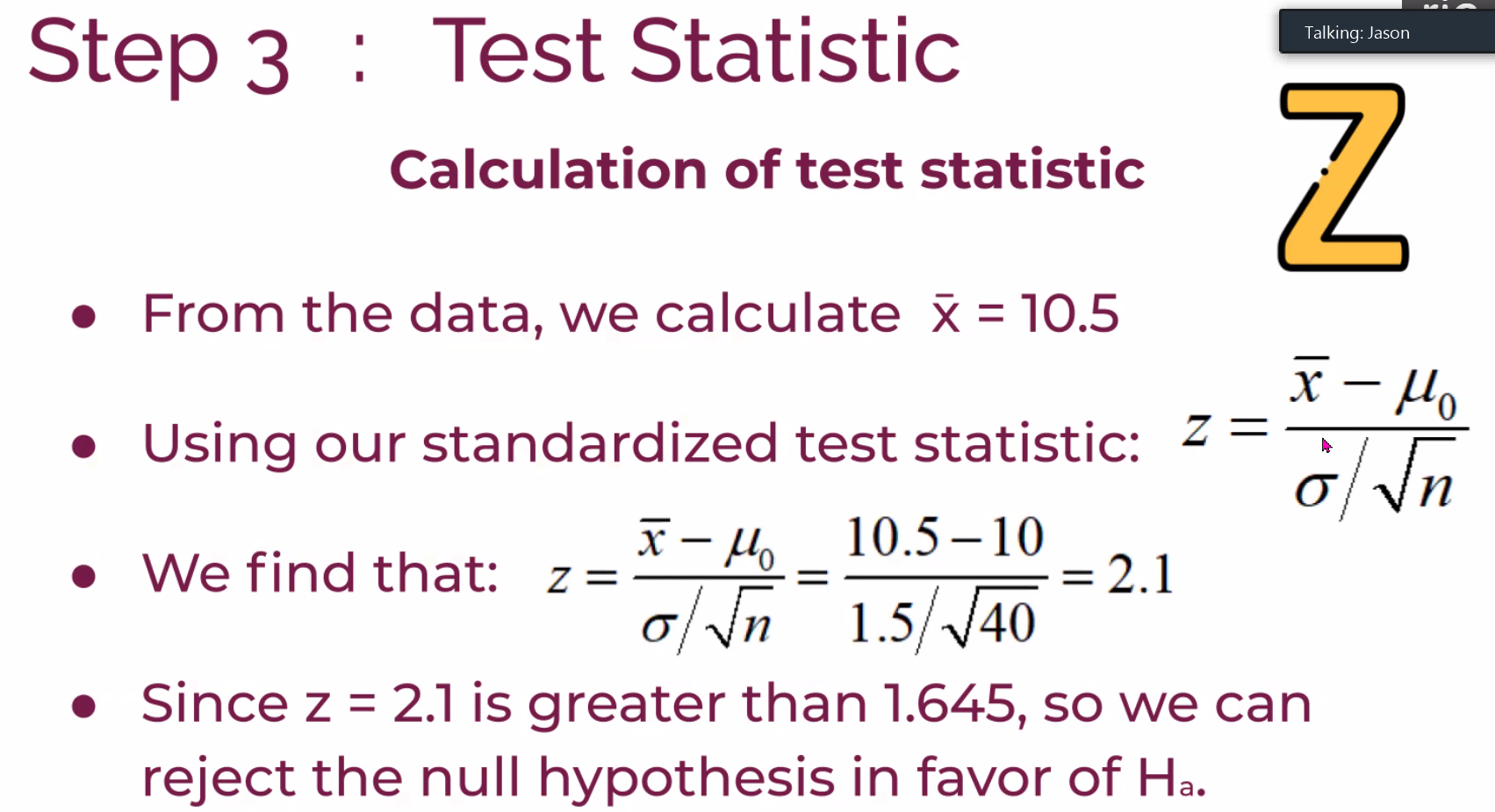
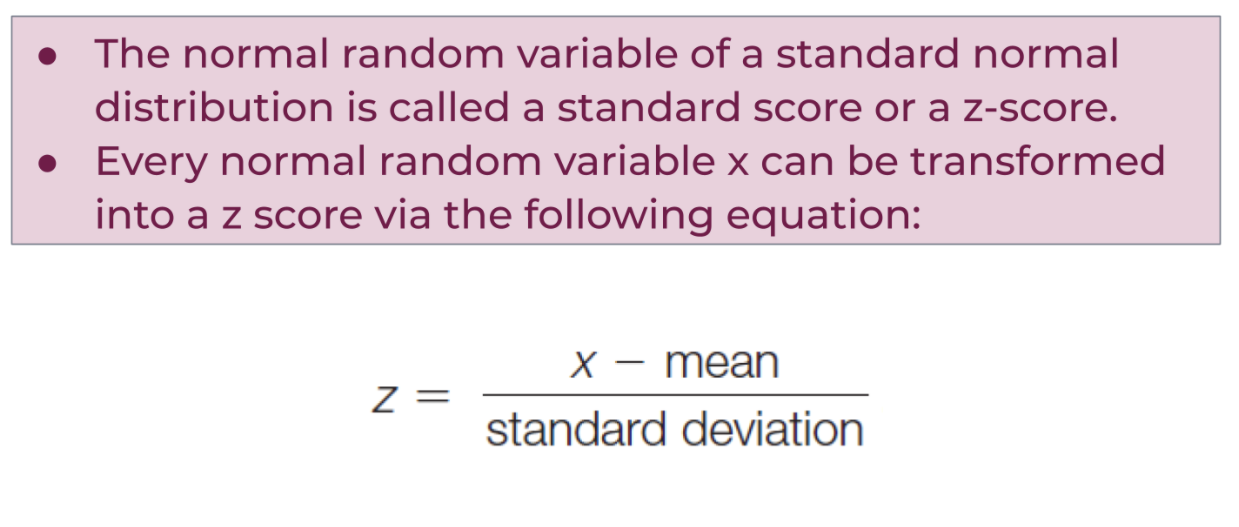
Ortalama değerimin 10 dan yüksek olduğunu ancak test istatistiğim sonucunda bulabilirim.



Mean= 10, Sample mean=10.5

10, 0’a denk geliyor, 10.5 da test istatistiğimin z sine denk geliyor. Burdan hareketle p-value ye ulaşacapım.

Test istatistiği formülüm:

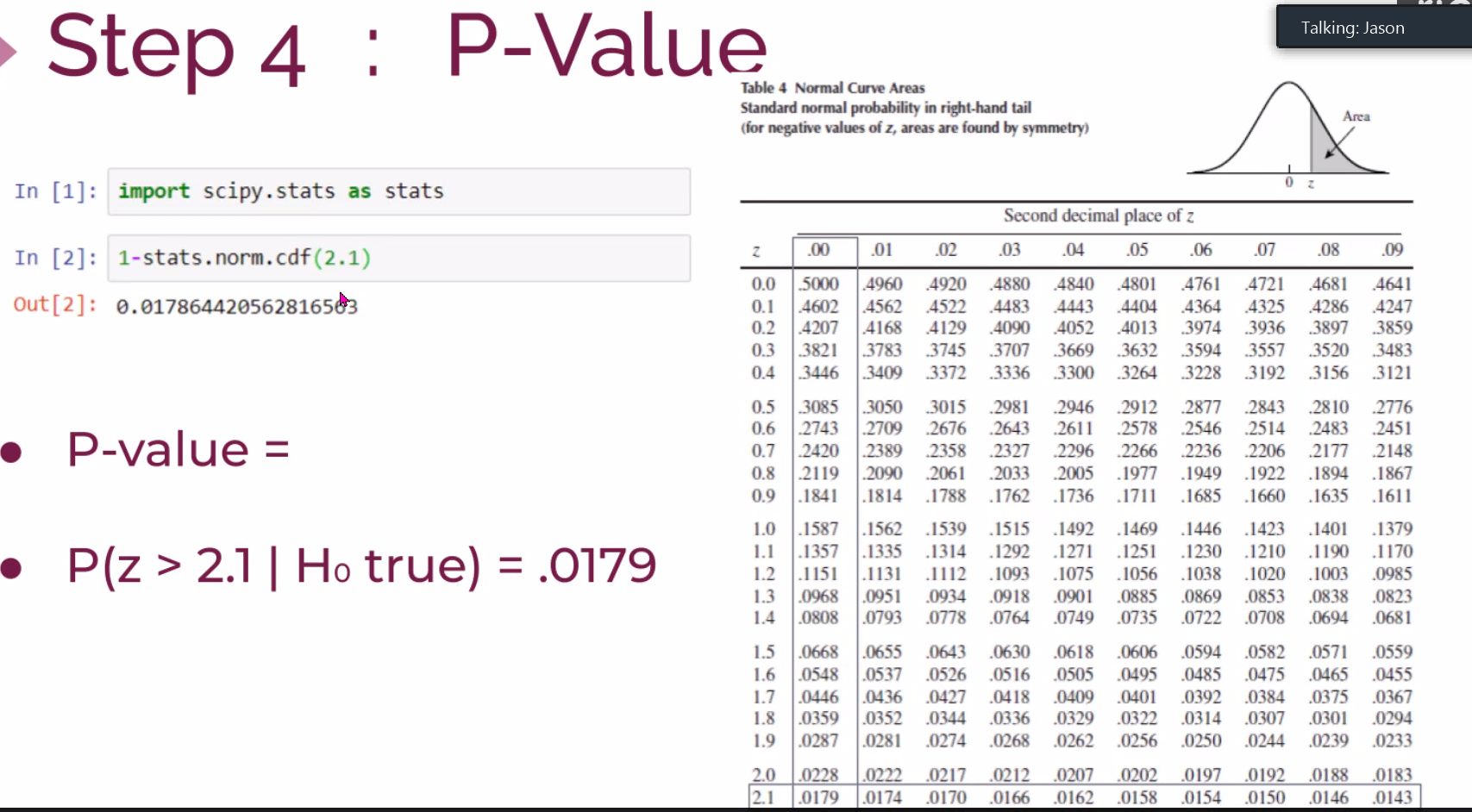
 

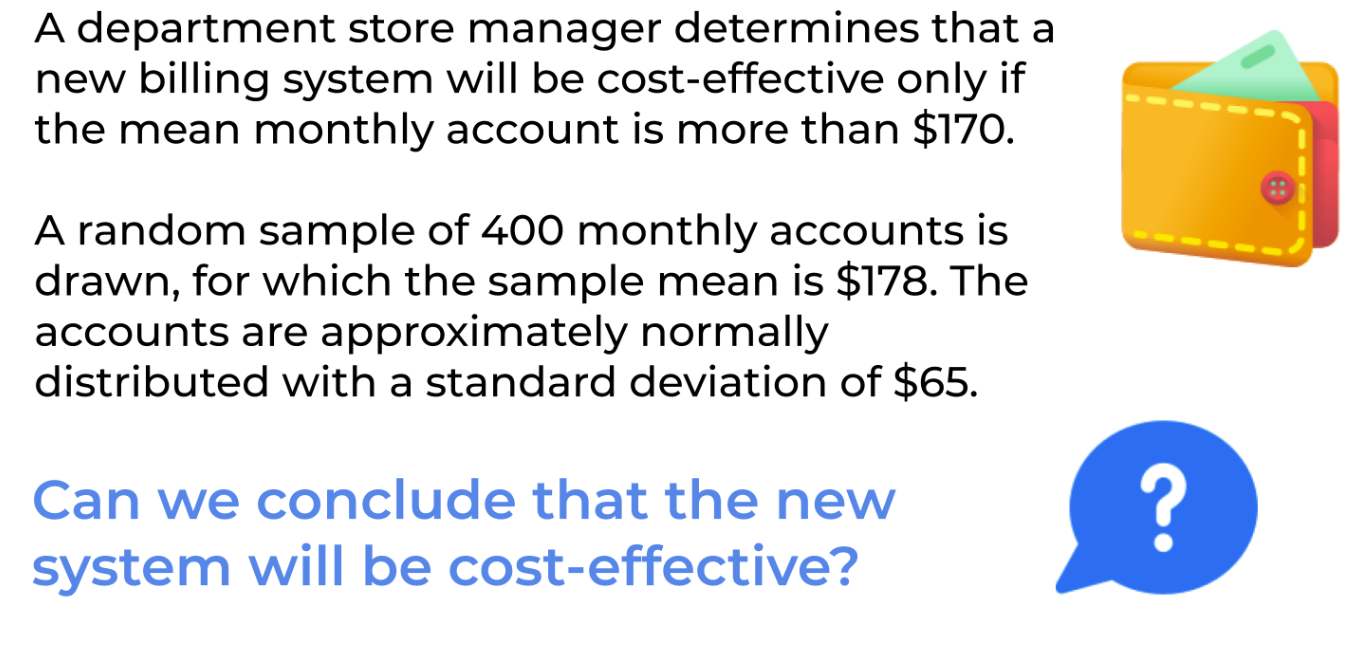
10.5 🡪 alfa :0.05 olduğu için??

Hesapladığım test istatistiğim : 2.1

Right hand tail tablosundan z değerine karşılık gelen değeri buluyorum. 0.0179 % bu p-value oluyor. (o z değerinin sağında kalan alan 0.0179 % lik alan)

P değeri test istatistiğinden (z) yola çıkarak hesaplanıyor ve daha önce verilen alfa ile kıyaslanıyor. Alfa verilmediği zaman 1’den Ci (güven aralığı) nı çıkartarak alfayı bulurum. Genelde Ci: 95 olduğunu kabul ediyoruz





Bir menager var. aylık 170 dolardan fazla ise yeni faturalandırma sisteminin cost efektif olduğu kabul edilecek.

Hesap bilgilerin normal dağılım o0lduğunu biliyoruz ve std, biliniyor (65)

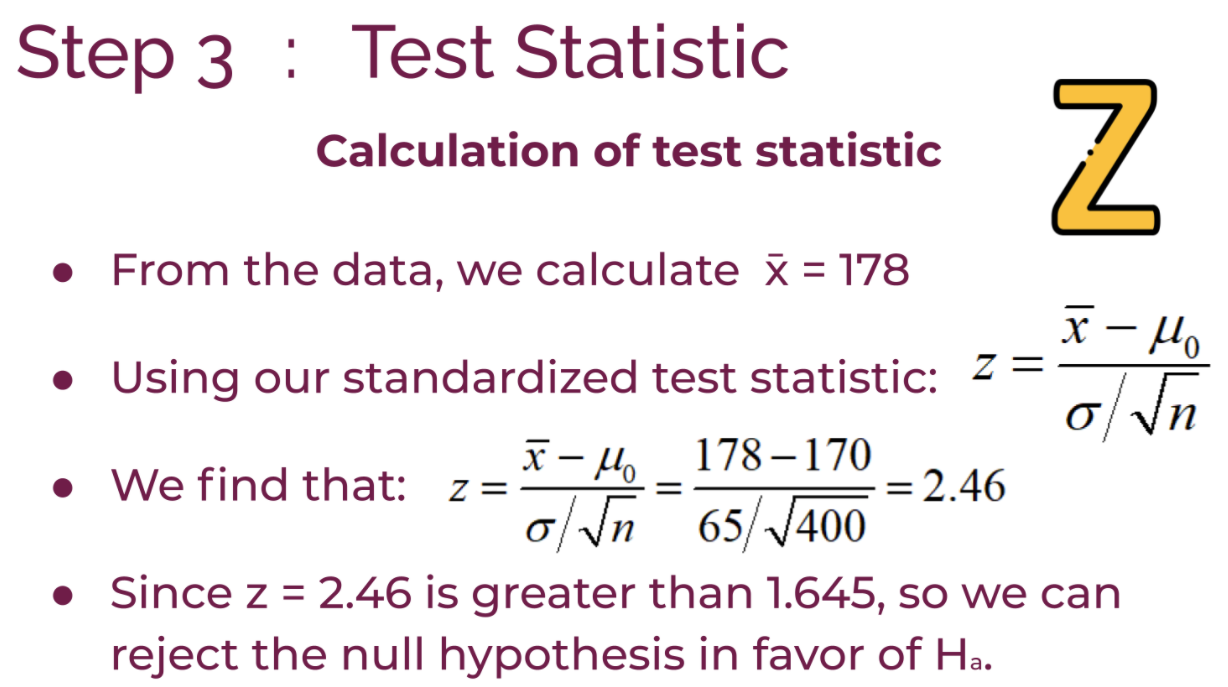
Sigma (std) biliniyor ve sample size 30 dan çok yüksek(400). Demek ki z dağılımını kullanabilirim.

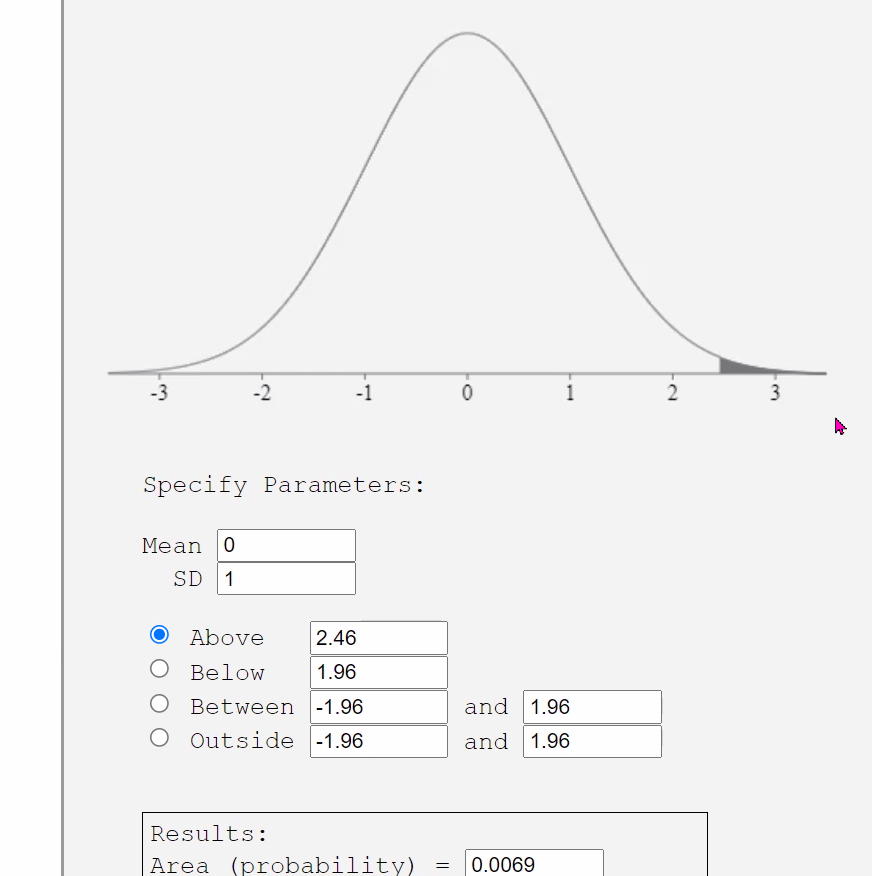
0 hipotez 🡪 mü = 170

Alternatif hipotez 🡪 mü > 170

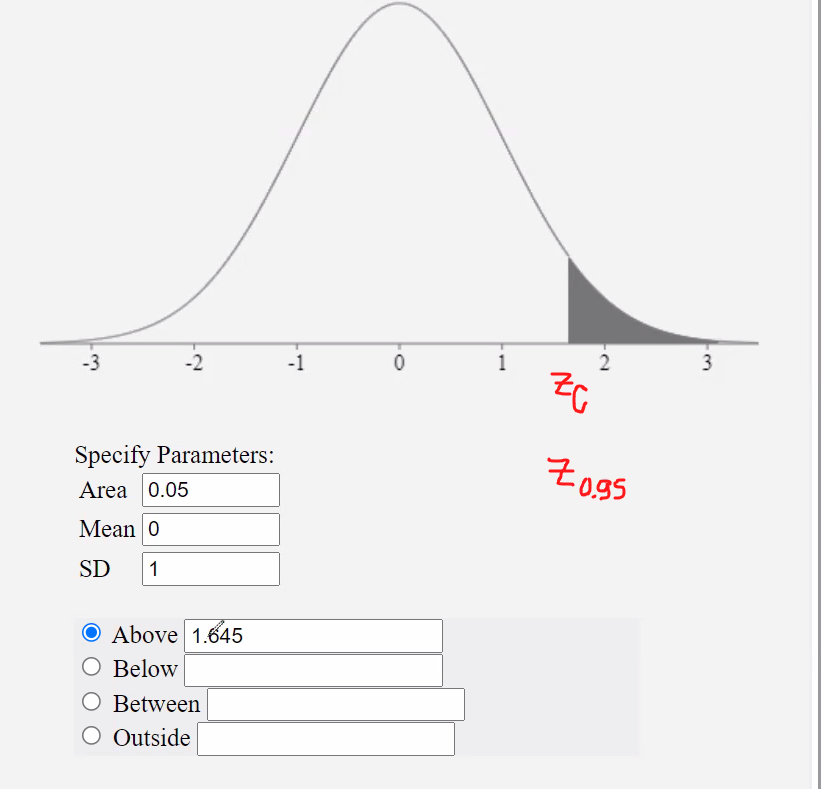
Büyüktür işareti var bu yüzden tek kuyruk kullanacağız.

Eşit değildir deseydi çift kuyruk kullanacaktık.

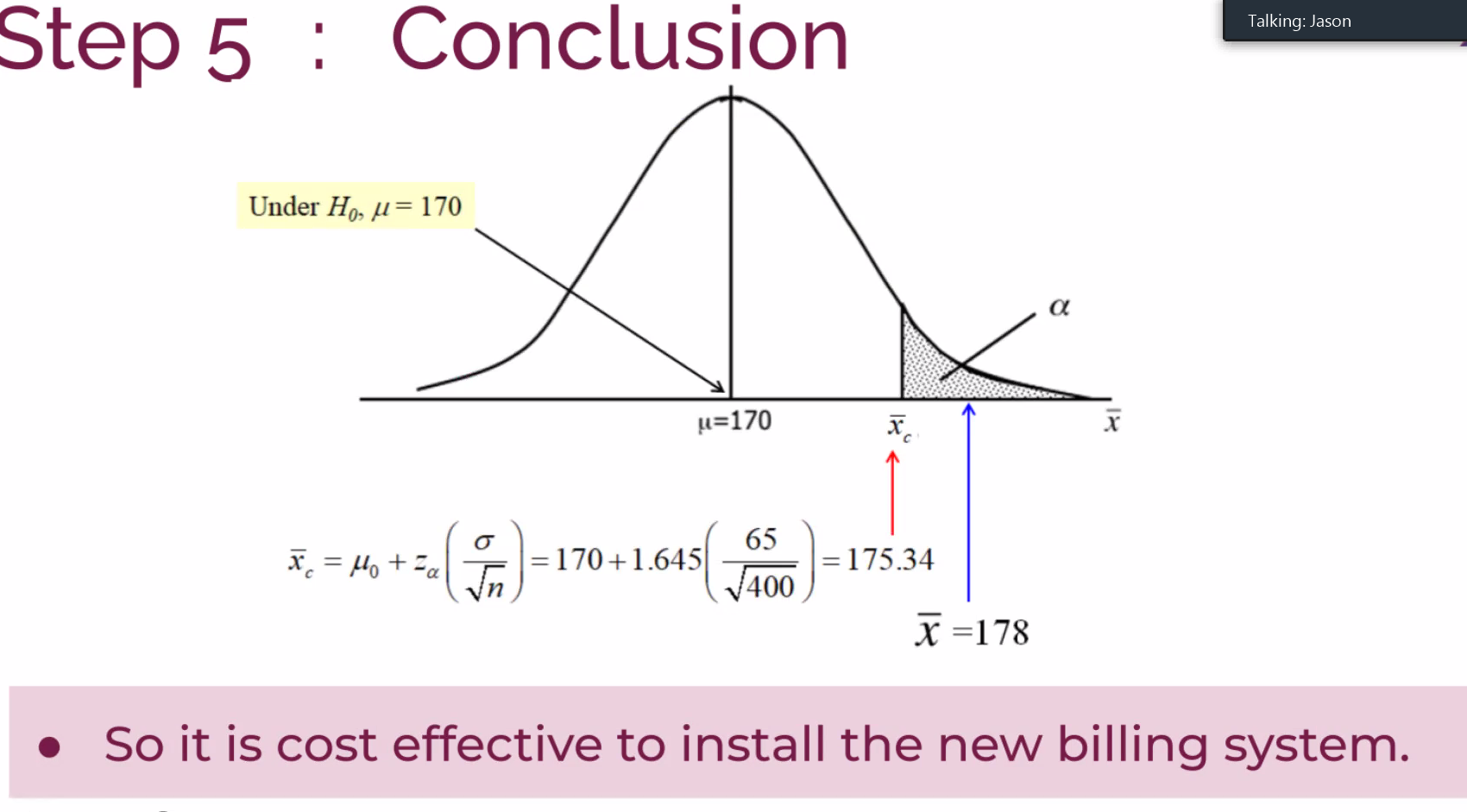




Z değerinin uçlarda çıkması (2.46), p değerinin küçük olması (0.0069 %) çok güçlü bir kanıt olduğunu gösteriyor.

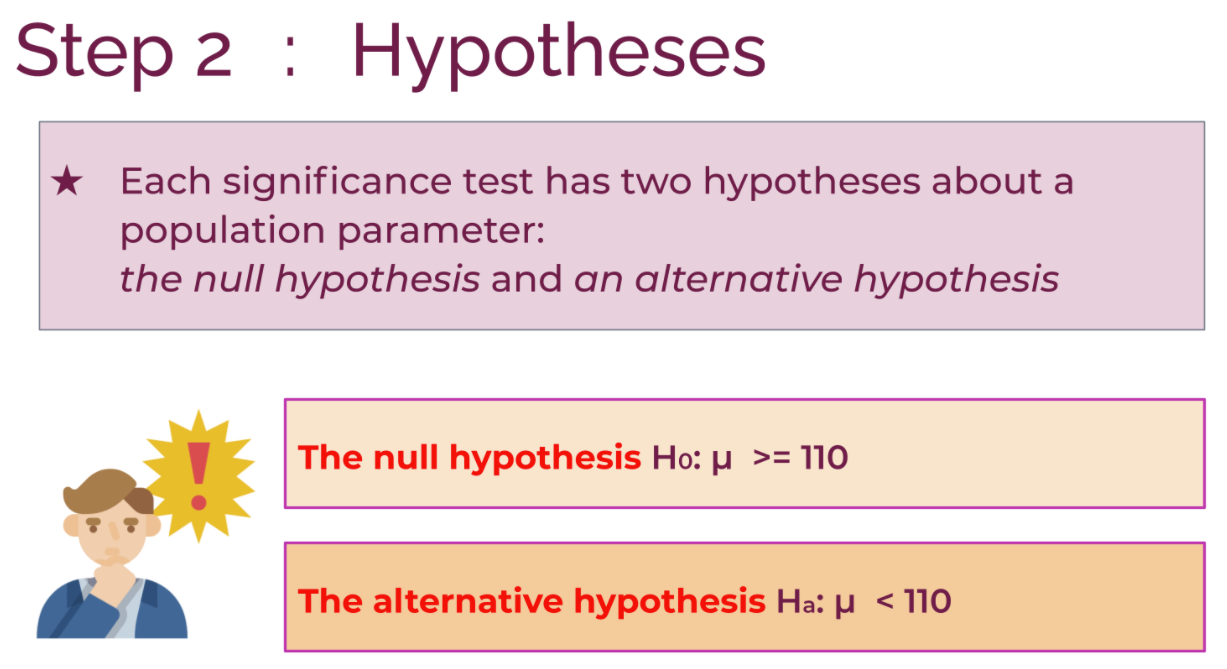


Kritik z değerim Zc, Ci’nin 95 olduğu değer. Test istatistiğimin z’si bu değerden büyük. Yani sonrasında kalan alan Zc nin alanından küçük.



Sample ortalama değeri 178 değerin taralı alana düştüğü için 0 hipotezini reddedebildim.

YENİ EXAMPLE. İLK YANSIYI KOY



Müdürün kabul ettiği mü >= 110

Alternatif hyptez mü <110

1 tail kullanacağız (<100 diyor)

HİPOTEZ TESTLERİ SAMPLING DISTRIBUTION DAN HAREKETLE ANLAŞILIYOR.

BUGÜNKÜ ÖRNEKLERİMİZDE ORTALAMA (MEAN) ÜZERİNDEN HİPOTEZ TESTLERİMİZİ YAPTIK.