

194. Hipotez Testi Nedir?

Bir manisi (bir soru, bir tahmini vs.) test etmek için kullanılan istatistiksel bir tekniktir.

Bir eczacı bir ilaç buldu diyelim. İlaç etkilidir gibi subjektif bir yaklaşımla ilaç piyasaya sunulamaz. İlaçın etkinliğini test edilmesi gerekir. Hem de öyle bir test ki ortaya çıkan etkinin sonsuza kadar sürmesi olasılığı göt önünde bulundurulduğu halde görmüş bir test olmalı.

İki örnek grubu var, bir gruba ilaç veriyoruz diğer gruba vermiyoruz. Her iki grubun da kan değerlerine bakıyoruz. Kanda hastalıklara ilgili bir belirtiye sahip olması lazım. Eğer ilaç gerçekten etkiliyse bunun kan ölçümlerine yansımaları gerekir fakat sonsuza kadar sürmesi de bir etki ortaya çıkmış olabilir. Hipotez testleri bu sonsuza kadar sürme durumunu da göt önünde bulundurarak sonuçta yer verilecek şekilde bize ilgilendirmiş olduğumuz konuda bir test etme imkanı sağlar.

Hipotez testinin önemli olmasının sebebi bize bilimsel bir dayanak noktası sağlayacak olmasıdır.

195. Hipotezler ve Türleri

Hipotezler null hypothesis (sıfır hipotezi) ve alternative hypothesis (alternatif hipotez) olmak üzere ikiye ayrılır.

H_0 : null hypothesis

H_1 : alternative hypothesis

Örneğin 10.000 götlemli bir ilaçtan 100 tane götlem bittimide gektik ve bunun öterinden ortalamaı hesapladık. Bunu hipotez testine uyarlayarak olursak hipotezimiz şu olabilir. Ana kottle ortalamaı, (population mean) buradaki yastarı ifade ettiğinde şöyle bir hipotez kurabiliriz. H_0 hipotezinde ana kottle ortalamaı yani bu ilaın yast ortalamaı 50'dir deditik sıfır hipotezinde H_1 'de ise 50'den farklıdır diyebiliriz. Daha sonra elimizdeki örnekle (100 götlemle) ortalamaın gerçekten 50 olup olmadığını sıayabiliriz. Bunu 2 yönlü deditirmiz $\mu \neq 50$ şeklinde ya da $\mu > 50$, $\mu < 50$ şeklinde üç farklı şekilde kurabiliriz.

$$H_0: \mu = 50$$

$$H_0: \mu \leq 50$$

$$H_0: \mu \geq 50$$

$$H_1: \mu \neq 50$$

$$H_1: \mu > 50$$

$$H_1: \mu < 50$$

19b. Hata Tipleri

Hipotez Testi Sonucu Verilen Karar

		H_0 reddedilmedi	H_0 reddedildi
Gerçek	H_0 doğru	Doğru Karar $(1-\alpha) \rightarrow$ Güven Düzeyi	I. Tip Hata α
	H_0 yanlış	II. Tip Hata β	Doğru Karar $(1-\beta) \rightarrow$ Testin Gücü

Hatalar; kurmuş olduğumuz hipotezleri test ettiğimizde / gerçekleştirdiğimizde ortaya çıkan sonuçlarla ilgili değerlendirmelerdir.

Populasyon ortasından (ilaaden) bir örnek çekmiştik 10.000 tane gıda değeri vardı. Bu bunların içerisinde 100 tane örnek çektik. Bunda yanlışlık olarak;

$$H_0: \mu = 50$$

$$H_1: \mu \neq 50$$

hipotezini kurmuş olalım. Yani ilaaden yaza ortalamaları 50'den diye düşünelim. Alternatif olarak ne kurduğumuzun bir örneği yok; 50 değildir, 50'den büyüktür ya da 50'den küçüktür şeklinde olsun.

Gerçek durumun ne olduğunu bilmiyoruz (populationı düşünsek). H_0 ifadesi doğru ya da yanlış olabilir. Ama bir test yapıyoruz belirlenmiş bir şekilde. Bir ana kitle 100.000 elimizde gibi söyledik ama bunlar gerçekte elimizde olmayacak. Dolayısıyla $H_0: \mu = 50$ hipotezinin doğru ya da yanlış olduğunu bilmiyoruz. Sol ekteki bu gerçek durumu ifade ediyor.

$$H_0: \mu = 50$$

$$H_1: \mu \neq 50$$

bunlara göre testimizi ^{bunu düşünce a priori yapıldığı} gerçekleştirdik diyalım. Yukarı ekteki karar bulma test sonucu bizim verdiğimiz karar olacak. Verdiğimiz karar kurmuş olduğumuz H_0 testini reddetmek veya reddetmemek şeklinde olacak.

• H_0 gerçekte doğru iken, $\mu = 50$ (göz ort = 50) eğer bunu reddedersek yanlış bir iş yapmış oluruz. Buna **I. Tip Hata / α hatası** denir.

• H_0 gerçekte doğru iken eğer H_0 'i reddetmemek doğru bir karar vermiş oluruz. Buna **güven düzeyi** denir ve $1-\alpha$ ile gösterilir.

• H_0 yanlış iken örneğin $H_0: \mu = 50$ dediğimizde gerçekte bu 50 değildi. Bilmiyorsa ama bunda hiçbir zaman da bilemeyeceğiz. Teknik bir dargınak bu. İşte bu H_0 gerçekte yanlış iken eğer bunu reddedemeysek yani yanlış olan şeyi kabul edersek bu durumda yine hata yapmış oluruz ve bunu **II. Tip Hata / β hatası** denir.

• Eğer H_0 gerçekte yanlış iken yani $\mu \neq 50$ doğru değil iken eğer bunu reddedersek doğru bir iş yapmış oluruz. Bu da **testin gücünü** gösterir ve $1-\beta$ şeklinde ifade edilir.

Daima bu hatalardan birini yapma ihtimali vardır. Bu ihtimaller risk derecesini belirleyerek ataltmak isteriz. Bir belirsizlik var ve bir belirsizlik ile ilgili yorumlar yapmaya çalışıyoruz.

Bizim genelde ilgilenerek olduğumuz **I. Tip Hata** dediyimiz α hatası.

✓ H_0 hipotezini kabul etmek gibi bir şey yoktur. H_0 hipotezini reddedemeyi veya reddedemeyi diye yorum yaparız. Çünkü gerçekte H_0 doğru iken onu reddettiğimizde yapacağımız hatayı, bilirken H_0 kabul ettiğimizde yapacağımız hatayı almıyoruz. (reddetmediklerinde $1-\alpha$)

Statquest with Josh Starmer - Hypothesis Testing and The Null Hypothesis
Bir hastalığı tedavi etmeye çalışıyoruz değilim. Bunun için elimizde ilaçlar var. Bu ilaçların hastaları tedavi etme süresini ölçüyoruz.

✓ A ilacını 3 hastaya veriyoruz, hepsi farklı farklı sürelerde iyileşiyor. B ilacını da aynı şekilde 3 hastaya veriyoruz. Hepsisi farklı farklı sürelerde iyileşiyor.

Her ilacın ortalamada ne kadar sürede hastaları iyileştirdiğine baktığımızda B ilacının iyileştirme süresinin A ilacının iyileştirme süresinden 15 saat fazla olduğunu görüyoruz.

Ve şöyle bir hipotez kuruyoruz:

"A ilacını kullanan kişilerin iyileşmek için B ilacını kullanan kişilere göre ortalama **15** saat daha az zaman ihtiyacı vardır."

Bunu başka hastalarda defalarca deneyip test edebiliriz.

• Bir taneinde A ilacını kullanan kişilerin iyileşme süresinin ortalamasının B ilacını kullanan kişilerin iyileşme süresinin ortalamasından **35** saat fazla olduğunu tespit ettik. (A, B'den fazla çıktı)

• Başka bir örnekte yine A, B'den fazla çıktı.

• A ve B'nin yanlış etiketlenmediğinden emin olmak için yine deney yaptık ve yine A, B'den fazla çıktı.
(Yanlış etiketlenmediğine zannat getirdik)

İlki dışında yaptığımız tüm deneyler hipotezimize ters sonuçlar üretti. Rahatlıkla ürettiği olduğumuzu hipotezi reddedebiliriz.

• İki tane daha ilacımız olduğunu düşünelim C ve D olmak üzere.

C ilacını 3 kişiye verdik diyetim ve D ilacını 3 kişiye verdik diyetim. C ilacını kullanan 3 kişinin iyileşme süresinin ortalamasının D ilacını kullanan 3 kişinin iyileşme süresinin ortalamasından 13 saat daha fazla olduğunu gördük.

Ve şöyle bir hipotez kurduk:

"C ilacını kullanan kişilerin iyileşmek için D ilacını kullanan kişilere göre ortalama 13 saat daha fazla zamana ihtiyaç vardır."

Bunu başka hastalarda test ediyoruz!

• Bir deneyimizde 12 saat fark görüyoruz (C, D den \checkmark)

• Başka bir deneyde 13,5 saat fark olduğunu görüyoruz.

A ve B ilaahorındaki yanlış etiketlenmenin burada olmadığını görüyoruz. Küçük farklar da kontrol edemeyeceğimiz random şeylerden kaynaklı.

Fakat yaptığımızda hipotezimizde 13 saat fark var demiydik, fakat denedikçe 12 veya 13,5 saat bulduk. Bu kadar küçük farklar hipotezimizi reddetmek için yeterli değil, fakat bu deneyler bizim hipotezimizin doğru olduğunu bir de aklı ikna etmiyor.

Tekrar eden deneylerden elde edilen sonuçlar hipotezimizi reddetmek için yeterli kadar farklı olmadığından aynı zamanda hipotezimizin doğru olduğuna dair de ikna etmediğinden yapabileceğimiz en iyi şey, hipotezi reddetmemektir. Çünkü hipotezimizi hangi deney sonuçlarına göre değerlendireceğimiz açık değil.

Biz hipotezimizi ilk deney sonucuna bakarak kurmuştuk ama mantıklı farklı hipotezler de kurabiliriz. 12,25 saat veya 13,1 saat daha fazla diyebiliriz. Hangisini test edeceğimizi nasıl bileceğiz?

• **Hedef** C ilacının D ilacından farklı olup olmadığını görmekse, hipotezi şu şekilde kurabiliriz.

"İlaç C ve ilaç D arasında iyileşme süreleri açısından bir fark yoktur."

Şeyler arasında hiçbir fark olmadığı hipotezine "Null Hypothesis" denir.

* Bu sefer E ve F ilaçları için deneyler yapıyoruz

E ve F arasındaki çok küçük farklar için $\mu_E = \mu_F$ hipotezini reddedemeyiz çünkü random küçük değişiklikler bu farklara sebep olabilir.

Farklı kişiler için deney tekrarladığımızda çok büyük farklar varsa rahatlıkla Null hypothesis'ımızı reddedebiliriz.

Eğer null hypothesis'ımız olmasa, deneyleri test etmek için başka deneylere ihtiyacı duyuyoruz. Bunun nedeni, bazı veriler elde edene kadar farkın 13 saat mı yoksa 13.000 saat mı olduğunu test etmemiz gerekip gerekmediğini bilmememizdir. Ama null hypothesis'da no difference 0'ı temsil eder.

Alternative Hypothesis: Main Ideas

Null hypothesis'ini reddedip reddedemediğimizi anlamak için verileri statistical test adı verilen bir şekilde analiz ediyoruz. Bu testin sonucunda Null hypothesis'ini reddedip reddedemediğimize karar vereceğiz.

İstatistiksel test 3 şeye ihtiyacı duyuyor:

1- veri

2- Null veya primary hypothesis (reddetmek için veya reddetmemek için)

3- Alternative hypothesis

alternative hipotez

Bu bağlamda basitçe C ve D ilaçları olan hastaların iyileşme sürelerinde farklılık var şeklinde belirledik yani null hypothesis'in tam tersi.

Alternative hypothesis'in istatistiksel testlerde kullanımını görecektir, bu testlerin nasıl çalıştığını göreceğiz. (Bu testler linear regression, t-test, anova, design matrices gibi testler.)

* İlk ilacın birlikte iyileşme sürelerinin ortalamasını alıyoruz ve her bir gözlemin bu ortalamaya olan uzaklığını hesap ediyoruz. Bunu her bir gözlemin C ve D ilaçlarının ayrı ayrı iyileşme süresi ortalamalarına olan uzaklıklarıyla karşılaştırıyoruz.

→ İlk bahsettiğimiz uzaklıklar Null Hypothesis'i temsil ediyor (C ve D'nin süre ortalamaları arasında fark yoktur)

İkinci bahsettiğimiz uzaklıklar Alternative Hypothesis'i temsil ediyor.

2 ayrı ortalamaya olan uzaklıklar, tek bir ortalamaya olan uzaklıklardan daha kısa ise verileri birleştirilerek aynı 2 farklı ortalamaya kullanmak daha mantıklı. Bu yüzden null hypothesisi reddedebiliriz.

Fakat verilerimiz böyle olur ki 2 ayrı ortalamaya olan uzaklıklar tek bir ortalamaya olan uzaklık arasında ciddi bir fark yoktur. O zaman iki farklı ortalamaya arasındaki fark aslında çok küçük kontrol edemeyişimiz random şeyler yansıtır. Tek bir hasta mı ağrıyı yapmıyordur o yüzden ağrıya jüresi uyanmıştır, belki sadece o ağrıyı yapmayı iki farklı ortalamaya da aynı olmaktadır.

Bu durumda null hypothesisini reddedemeyebiliriz.

• Sadece iki grup verisi olduğunda alternative hypothesisin ne olduğu çok açık, çünkü direkt null hypothesisin tam tersi.

Fakat 3 veya daha fazla grup için bakıldığında alternative hypothesis daha farklı değerlendirilmelidir.

Aynı şekilde null hypothesisimiz: There is no difference in recovery times between Drugs C, D and E olarak.

Fakat alternative hypothesis için farklı seçeneklerimiz olacak:

- C, D ve E ilaçlarının hepsi birbirinden farklıdır. Bu durumda her ilacın ortalamasıyla her bir hastanın arasındaki uzaklığı hesaplayacağız.

- C ve D arasında fark yoktur, E ayrı tutuluyor. Bu durumda C ve D için tek bir ortalamaya hesaplayacak ve her bir hastanın arasındaki farkı bakacağız. E'nin kendi ortalamasını hesaplayacağız ve her hastanın bu ortalamaya olan uzaklığını hesaplayacağız.

İstatistiksel testle bu hipotezlerden hangisini kullandığımızla ilgili olarak Null Hypothesis hakkında farklı bir karar verebiliriz. Bu yüzden hangi alternatif hipotezi doğru kullanmak istediğimizi açıkça belirtmek önemlidir. Ancak kullandığımız alternatif hipotezden bağımsız olarak yalnızca primary veya Null Hypothesisi reddedebilir veya reddedemeyiz.

İkinci söylediğimiz alternatif hipotezi kullanarak null hypothesisi test edersek ve null hypothesisi reddederek bu alternatif hypothesisi lehine reddettiğimizi söyleyebiliriz. Fakat bu alternatif hipotezi kabul ettiğimizi söyleyemeyiz, çünkü başka alternatif hipotezler daha iyi olabilir. Başka bir deyişle, doğru olanı kabul edip etmediğimizi test etmek için çok fazla olasılık var.

Özetleyecek olursak;

En basit istatistiksel test için 3 seçeneğimiz var demektir: Veri, null veya primary hypothesis ve bir alternatif hypothesis.

- Eğer sadece iki grup verisi varsa alternatif hipotezin ne olduğu bellidir.
- Eğer 3 veya daha fazla grubumuz varsa alternatif hypothesisin ne olduğunu seçeriz. İstatistiksel test sırasında hangisini kullandığımızla ilgili olarak, null hypothesis hakkında farklı bir karar verebiliriz.