

HCI'da Deneysel Arařtırmalar
(Experimental Researches)
İncelemesinin Devamı
(Ders Notu 3 Devamı)

Önem Testi (Significance) Testi Niçin Önemlidir?

- ❑ Hemen hemen tüm deneysel araştırmalar önem (significance) testleri ile analiz edilir ve raporlanır.
- ❑ Eğer rasgele olarak tasarlanmış bir arayüz çalışmasında gazete makaleleri ya da konferans çalışmaları inceleniyorsa, aşağıdakilere benzer ifadelerle karşılaşılabilir:
 - «Katılımcılar önem testinde dinamik açıdan statik duruma göre ortalama olarak daha iyi performans göstermişlerdir. $F(1,25) = 20.83, p < 0.01$ değerleri çıkarılmıştır.»
- ❑ Bir başka örnek için ise:
 - «Girişlerin dokunmatik (tactile) olarak yapıldığı durumda girilen satır sayısında önemli farkın olduğu bir test için $(t(11) = 6.28, p < 0.001)$ değerleri elde edilmiştir.»

İstatistiksel Önem ne Zaman Önemli Değildir?

When is statistical significance not significant?

Ortalama olarak öğrencilerin okulda geçirdikleri yıllar ile hırsızlık yapmaları arasında ortalama -0.7999 değerinde **false –negatif** korelasyonu değerinde bir sonuç olsun.

- Bu ilişki istatistiksel olarak (**p değeri $<0,000$**) önemlidir. 113 örnek ülkedeki deneylerden elde edilmiştir.
- Öğrencilerin ortalama olarak okulda geçirdikleri yılları ile hırsızlıkları arasında **negatif** bir ilişki olduğu sonucu çıkarılır.

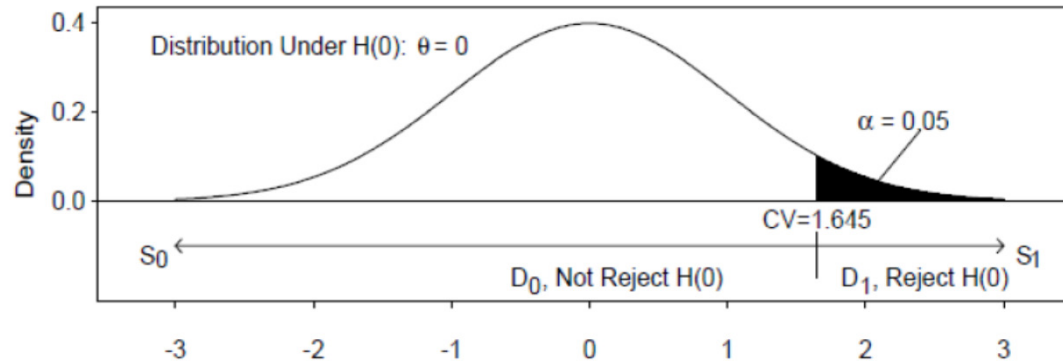
Diğer bir ifade ile:

H0 boş (null) hipotezi böyle bir ilişki olduğunu reddederken

H1 alternatif hipotezi böyle bir ilişkinin var olduğunu savunur.

İstatistiksel Önem ne Zaman Önemli Değildir?

When is statistical significance not significant?



- ❑ Test istatistiğinin değeri reddedilen bölgeye düştüğünde, **boş hipotez** alternatif hipotez lehine olacak şekilde **reddedilir**.
- ❑ Aksi halde boş hipotezi reddetmek için yeterli delil bulunmadığı gerekçesiyle, boş hipotez geçerli olur.”
- ❑ Boş hipotezin reddedildiği seviye genellikle 100 denemeden 5 veya daha az sayıda gerçekleşme olarak belirlenir.
 - ❖ Böyle bir ilişki 100 denemeden 5’den az meydana gelmesi durumunda önemli değildir anlamındadır.

Type I ve Tip 2 Hataları (Errors)

- ❑ **Önem testi** (significance testing) bir boş (null) hipotezinin (H_0) karşıtı olan bir alternatif bir hipotezdir (H_1).
- ❑ **Önem testi** ile boş (null) hipotezin doğruluđu olasılıđı belirlenir.
- ❑ Önem testlerinin tümü Tip1 ve Tip2 hatalarının risklerine maruz kalır.
- ❑ Tip1 hatası, α hatası (error) ya da “**false positive**” olarak adlandırılır.
 - ❖ Boş (null) hipotez doğru olduđuunda ve reddedilemediđinde, Tip1 hatası boş hipotezin ret edilmesinin hatasını (mistake) ifade eder.
- ❑ Boş (null) hipotez yanlış (false) olduđuunda ve reddedilebildiđinde, Tip 2 hatası β hatası (error) ya da “**false negative**” olarak adlandırılır.

Önem Testi : Genel bir Örnek

Aşağıdaki boş (null) ve alternatif hipotezler verilsin:

H_0 : Davalı suçsuzdur. H_1 : Davalı suçludur.

□ Type I hatası jürinin davalının suçlu olduğuna karar vermesi ile oluşur.

❖ Tip 1 hatasının anlamı, davalının gerçekten suçsuz olması, fakat boş (null) hipotezin doğru olmasına rağmen reddedilmesi anlamındadır.

□ Tüm tip 2 hataları davalı gerçekten suçlu olduğunda jürinin davalının suçsuz olduğuna karar vermesi ile sonuçlanır.

❖ Tip 2 hatasının anlamı, boş (null) hipotezin gerçekte yanlış (false) olması durumunda kabul edilmesidir.

Önem Testi : Genel bir Örnek

	Jüri Kararı: Suçsuz	Jüri Kararı: Suçlu
Gerçek: Suçsuz	No Error	Tip 1 Hata (Error)
Gerçek:Suçlu	Tip 2 Hata (Error)	No Error

- İdeal durumda (gerçekleşmesi istenilen) davalının gerçekten suçlu olması durumunda jürinin suçlu olduğu kararı vermesidir; ya da gerçekten suçsuz ise suçsuz olduğu kararını vermesidir.
- Fakat gerçek böyle değildir: Jüri seyrek te olsa hatalar yapabilir (mistake) .
- Her iki tip hatanın da bir maliyeti vardır.

HCI Örneđi

Bir bankada bir HCI arařtırması olarak , dokunmatik ekran arayüzlü ATM'lerin, banka řubelerindeki basmalı (düğmeli) ATM'lerden daha kolay kullanıp kullanamayacakları arařtırılsın.

H0: Boř hipotez: ATM 'lerdeki dokunmatik ekranlarla düğmeli ekranlar arasında fark yoktur.

H1:alternatif hipotez: ATM 'lerdeki dokunmatik ekranları kullanmak düğmeli ekranları kullanmaktan daha kolaydır.

Type I and Type II Errors in a Hypothetical HCI Experiment

	Study Conclusion: No Difference	Study Conclusion: Touchscreen ATM is Easier to Use
Reality: No Difference	No error	Type I error
Reality: Touchscreen ATM is Easier to Use	Type II error	No error

HCI Örneği

Tip 1 hataları Tip2 hatalarından daha kötüdür.

- ❑ Araştırma sonucunda dokunmatik ekranları kullanmanın düğmeli ekranları kullanmaktan daha kolay olduğuna karar verilirse , fakat bu gerçekten doğru olmadığında Tip 1 hatası gerçekleşir.
- ❑ Tip 2 hatası oluştuğunda dokunmatik ATM ekranlarının basmalı (butonlu) ekranlardan daha iyi olmadığı, her iki şekilde etkileşimde bir fark olmadığı kararı verildiğinde gerçekleşir. Fakat gerçek bu değildir; dokunmatik ekranları kullanmak daha kolaydır.

Önem Testinde p değerinin önemi

- ❑ P değerleri Tip1 ve Tip2 hatalarının kontrolünde kullanılır.
- ❑ Çok küçük bir P değeri (0.05) Tip1 hatalarının oluşumunun kontrolünde yaygın olarak kullanılır.
- ❑ Eğer bir önem testi $P < 0.05$ önemde bir değer döndürürse Tip 1 hatası yapmanın olasılığı 0.05 in altındadır.
 - ❖ Bu şöyle ifade edilir. Boş hipotezi yanlış bir kararla reddetmenin olasılığı 0.05 den küçüktür.
- ❑ Tip 2 hatalarını belirlemek ve P değerini azaltmak için, örnek büyüklüğünün Tip 1 hatasının ölçülmesi kriterine göre daha fazla sayıda uygulamadan seçilerek daha büyük olması gerekir.

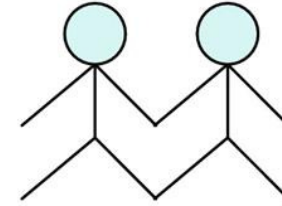
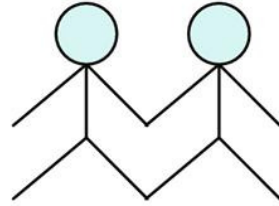
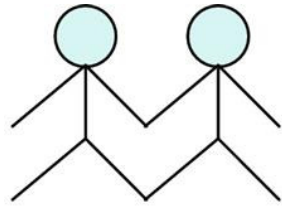
Örnek bir Deneysel Tasarım

- Deneysel bir tasarımın hipotezi: Farklı boyutlardaki (küçük, orta ve büyük) simgeleri seçmek için fare, joystick veya trackball kullanırken hedef seçim hızı arasında fark yoktur.
- Bu hipotezin **2 bağımsız değişkeni** vardır: işaretleme aygıtının türü ve simge boyutu (type of pointing device and the size of icon).
- 3 farklı işaretleme aygıtı test edilecektir : mouse, joystick ve trackball
 - ❖ **Bağımsız değişkenler** için kullanılacak 3 **koşul (condition)** olduğu vurgulanmıştır.
- Test edilecek 3 farklı hedef büyüklüğü vardır: small, medium ve large.
 - ❖ Bu **bağımsız değişken** için de önerilen 3 **koşul (condition)** vardır
- İki bağımsız değişkenin her bir kombinasyonunu test etmek için, iki bağımsız değişken sonucunun her bir deney için toplamda $3*3=9$ koşulda birleştirilmesi demektir.

HCI Deneylerinde Wizard of Oz Önemi

- ❑ Kullanıcının söylediklerini dinleyerek kullanıcının yorumlarının sisteme girişini gerçekleştiren bir «speech recognizer» olarak bir insan gibi davranır.
- ❑ Katılımcılara gerçeğin ne olduğu, elde edilmek istenilen deney tamamlanana kadar açıklanmaz.
- ❑ Wizard-of-Oz yaklaşımı gerçekte mevcut olmayan ama gerçekleşmesi hedeflenen (ideal) uygulamaların test edilmesini sağlar.
- ❑ İnsanlar da hata yapabildiği için «wizard» ın da konuşmayı dinlerken ya da kelimeleri sisteme aktarırken hata yapma olasılığı vardır.
- ❑ Bu nedenle de bağımsız değişkenlerin gerçekleştirilmesi istenilen koşulların kontrol edilmesi oldukça güçtür.

Gruplar Arası Tasarım (Between Group Design)



Testi gerçekleştirecek kullanıcılar bağımsız değişkenlere göre gruplandırılmıştır.



QWERTY
keyboard

DVORAK
keyboard

Alphabetic
keyboard

Grup ii Tasarım (Within Group Design)

Testi gerekleřtirecek kullanıcılar bağımsız deęişkenlerin tümünü test eder.

